NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
United States Department of Commerce

HURRICANE FORECAST IMPROVEMENT PROJECT

2020 HFIP R&D Activities Summary: Recent Results and Operational Implementation

May 2021

HFIP Technical Report: HFIP2021-1

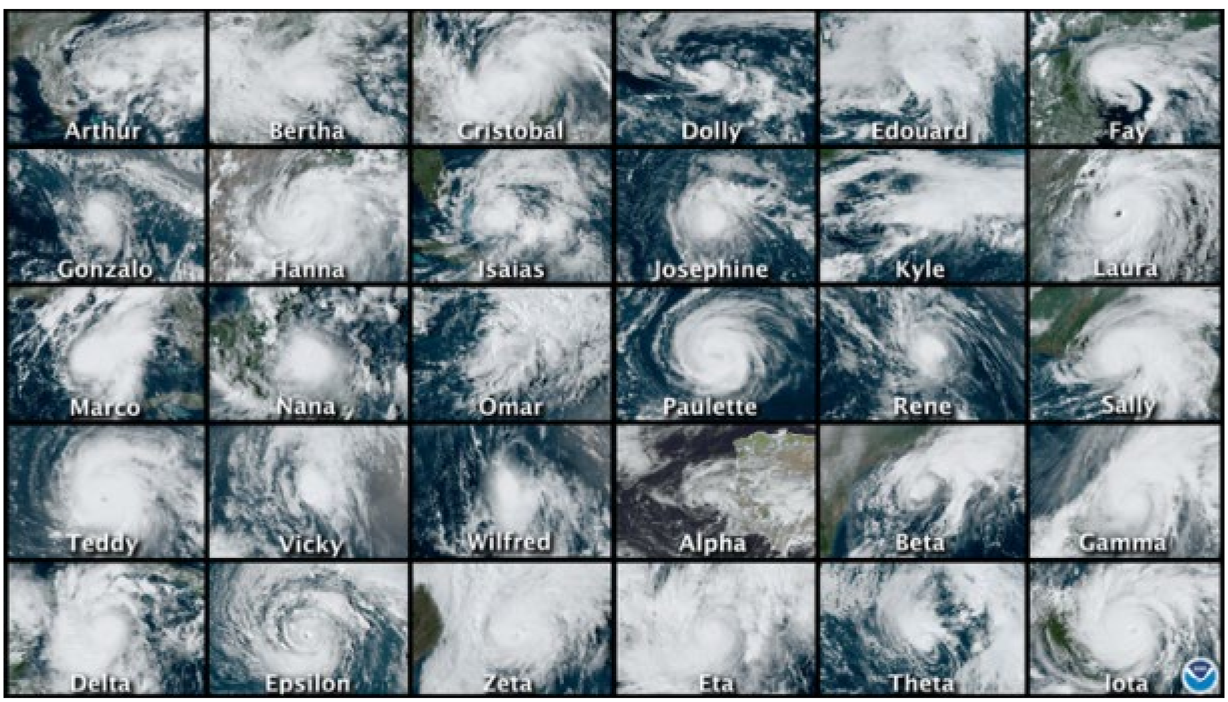

Image on the cover page shows an extremely active 2020 Atlantic hurricane season with a record-breaking thirty named tropical storms.

2020 HFIP R&D Activities Summary: Recent Results and Operational Implementation

S. Gopalakrishnan[11], S. Upadhayay[7,18], Y. Jung[7], F. Marks[11], V. Tallapragada[8], A. Mehra[8], M. Brennan[9], M. DeMaria[1,9], A. Aksoy[11,22], G. Alaka Jr.[11], C. Alexander[13], M. Bender[17], L. Bernardet[2,13], M. Biswas[2,6], J. Cangialosi[9], J. Dong[3,8], R. Dunlap[5], M. Ek[2,6], J. L. Franklin[9,19], L.Gramer[11,22], L. Harris[15], A. Hazelton[11,22], J. S. Hilderbrand[10], E. Kalina[2,13,21], H.-S. Kim[12], D. Koch[16], P. Kucera[6], B. Liu[3,8], P. McCaslin[13], T. Marchok[15], J. Moskaitis[4], K. Musgrave[1], L. Nance[2,13], K. Newman[2,6], M. Onderlinde[9], A. Penny[9], W. Ramstrom[11,22], E. N. Rappaport[9], J. Sims[7], J. Sippel[11], R. Torn[20], X. Wang[23], W. Wang[3,8], Y. Weng[3,8], J. Whitaker[14], H. Winterbottom[14,21], B. C. Zachry[9], D. A. Zelinsky[9], C. Zhang[3,8], X. Zhang[11], Z. Zhang[8], and L. Zhu[3,8].

May 2020

[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 of Colorado, CIRES, Boulder, CO
[22] University of Miami, CIMAS, Miami, FL
[23] University of Oklahoma, Norman, OK

This page is intentionally left blank.

Table of Contents

This page is intentionally left blank.

List of Figures

List of Tables

Executive Summary

This technical report describes the activities and results of the Hurricane Forecast Improvement Program (HFIP) that occurred in 2020. The major development focus remained in building the next generation hurricane model, the Hurricane Analysis and Forecast System (HAFS) within the Unified Forecast System (UFS), primarily for track and intensity predictions. This report summarizes the progress made in 2020, including model developments and the second year of progress made towards transitioning the next generation hurricane modeling system, HAFS, into operations.

The 2020 Atlantic hurricane season was the busiest season, ever, on records. There were 30 named storms, of which 14 developed into hurricanes, with 7 of those becoming major hurricanes. In the east Pacific, there were 17 named storms, of which 4 developed into hurricanes with 3 major hurricanes. There were 31 rapid intensification (RI) events in a total of 510 forecasts issued in the Atlantic. There were 11 landfalls in the U.S. with four landfalls in Louisiana alone.

The major highlights of 2020 were:

1. The upgraded version of the Hurricane Weather and Research Forecasting (HWRF) model was implemented on July 22, 2020. The major upgrades were two-way ocean coupling, one-way wave-model coupling and the high-resolution land-sea masks for the moving nests.
2. In the Atlantic Basin, HWRF had the best intensity skill of any individual model at all lead times. HWRF performance was outstanding, especially in the Gulf of Mexico. In the northeast Pacific basin, HWRF was a strong performer through day 4, but intensity skill dropped sharply after that.
3. HWRF had reasonable track skill in the Atlantic basin, but lagged Global Forecast System (GFS) at all times. The model did not perform very well in the east Pacific basin either.
4. In the Atlantic basin, intensity skill for the Hurricanes in a Multi-scale Ocean-coupled Non-hydrostatic (HMON) model was not as good as HWRF. However, HMON had slightly better track skill than HWRF. In the eastern Pacific basin, HMON intensity skill was slightly better than HWRF at 48 and 60 h but degraded after that. HMON track skill had a good start at 12 and 24 h but this didn’t continue at longer forecast lead times.
5. The basin-scale HWRF (HWRF-B) was run in real-time in parallel along with HWRF in the operational machine (WCOSS). HWRF-B showed improvement in intensity forecasts over the operational HWRF, while track skill was much lower (negative) compared to HWRF. The HWRF-B continues to show value for multiple high-resolution nests within a single outer domain.
6. Comparison of the HFIP rapid intensification (RI) performance metric for 2018-20 against the 2015-17 baselines is encouraging. At 24 hour the baseline error was reduced by 27%, at 48 h the baseline error was reduced by 19%, and at 72 hour the baseline error was reduced by 23%.
7. Another major accomplishment in 2020 was the continuing development of NOAA’s next-generation HAFS. The regional HAFS was coupled with the HYbrid Coordinate Ocean Model (HYCOM) to account for air-sea interaction. Some significant progress was made with the development of the moving nest in the global version of HAFS, and this development is being transitioned to the regional version for operational implementation in 2023.
8. Four configurations of the HAFS model were run as part of the 2020 HFIP Real-time Experiment (HREx). The four configurations were (i) the ocean coupled, high resolution regional model (HAFS v0.1A); (ii) global model with a high resolution nest (HAFS v0.1B); (iii) uniform grid projection on Extended Schmidt Gnomonic (ESG) grid (HAFS v0.1J); and (iv) HAFS ensembles with 18 members (HAFS v0.1E). All HAFS configurations did well with track

skill in the Atlantic basin with HAFS v0.1B showing exceptional track prediction skill. The global nested version of HAFS performed better than HWRF and GFS at all lead-times beyond 24 hours and at least 10-20% better than HWRF at those lead times.

9. The hurricane supplemental enabled some accelerated initial developments of HAFS. These 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 into the UFS developments.

10. Under the 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 activities, and gusts, as well as sustained winds) at actionable lead times for emergency managers (e.g., 72 hours). While significant progress was made in 2020, especially for track and intensity predictions using the HWRF system, further improvements are necessary. The HAFS system is expected to address these new HFIP goals.

1. Introduction

This report describes the Hurricane Forecast Improvement Program (HFIP), it's 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. The first part of this report describes the background of the program. This year's report focuses upon capturing state-of-the-art HFIP modeling accomplishments during the 2020 hurricane season, continued development of the Hurricane Analysis and Forecasting System (HAFS) within the Unified Forecast System (UFS) and future plans. For more background information, readers are referred to earlier reports available on the HFIP website.

The 2020 Atlantic hurricane season was the busiest year since 2005 and second busiest since 1990. There were 30 named storms, of which 14 developed into hurricanes (*cover page image*), with 7 major hurricanes. In the east Pacific, there were 17 named storms, of which 4 developed into hurricanes with 3 major hurricanes. There were 31 occurrences of Rapid Intensification (RI)[1] events in the Atlantic basin in the 510 forecasts issued, and 26 in the East pacific basin out of 407 issued forecasts. Three hurricanes in 2020, Hanna, Laura and Zeta, all rapidly intensified before making landfall in the United States. There were 11 landfalls in the U.S. with 4 landfalls in Louisiana alone.

2. The Hurricane Forecast Improvement Program (HFIP)

Twenty-seven named tropical storms and thirteen hurricanes crossed the U.S. coastlines from 2000-2010. HFIP was established within NOAA in June 2007, in response to particularly damaging hurricanes (e.g., Charley, 2004; Wilma, Katrina, Rita, 2005) in the first half of that decade. HFIP's 5-year (for 2014) and 10-year goals (for 2019) are:[2]

- 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)[3] for RI to 90% at Day 1, decreasing linearly to 60% at day 5, and decrease 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. Through HFIP, NOAA continues to improve the accuracy of hurricane forecasts, with applied research using advanced computer models.

---

[1] RI for hurricanes is defined as an increase in wind speed ≥30 knots in 24 hours. The HFIP goal also applies to rapid weakening (RW) of a decrease of 25 knots in 24 hours.

[2] 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.

[3] 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 but should have forecasted RI). 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. Hurricane Forecast Improvement Program’, instructing 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, with a 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 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 will be a multi-scale model and data assimilation package capable of providing analyses and forecasts of the inner core structure of the TC out to 7 days, which is key to improving size and intensity predictions, as well as the large-scale environment that is known to influence the TC's motion. 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 (DA), 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; and (iii) 3A-2: Accelerate Re-engineering of HAFS.

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 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 Laboratory (ESRL) in Boulder, Colorado. For instance, in the 2020 season, an advanced version of Hurricane Weather and Research Forecasting (HWRF) model, called the Basin-Scale HWRF, and four versions of HAFS were tested in near real-time within the stream-2 HREx. *(See section 9 for results)*

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 Atlantic basin. For track, the ensemble members were the operational aids GFSI, GFDI, UKMI, NGPI, GFNI, and EMXI (GFS, GFDL, UKMET, NOGAPS, new GFDL and ECMWF model with 6-hour interpolation respectively), while for intensity the members were GFDI, DSHP (Decay Statistical Hurricane Intensity Prediction Scheme model), and LGEM ((Logistics Growth Equation Model)[4] (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 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.

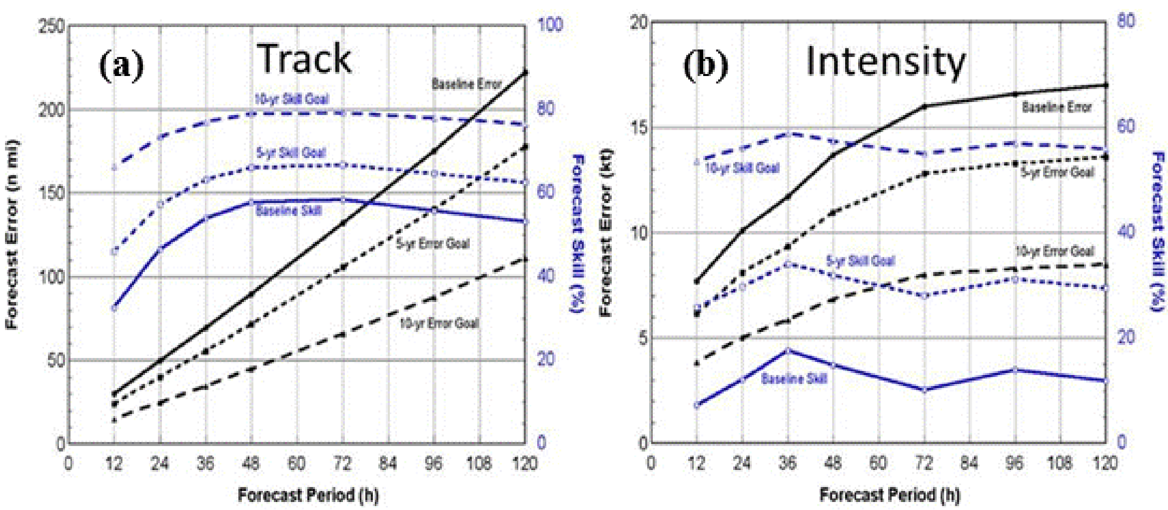

Figure 1. (a) Track and (b) Intensity Error Baseline and Goals, where the forecast errors are represented by black lines labeled on the left side of the graph, and the forecast skill is represented by blue lines labeled on the right side of the graph. Solid black lines represent baseline forecast errors, 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.

[4] See Appendix for details on operational aids (GFSI, GFDI, UKMI, NGPI, GFNI, EMXI,GFDI, DSHP, LGEM)

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.

4. 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 since the 60'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 data assimilation (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. However, TCs like Sandy (2012), Joaquin (2015), Florence (2017) and Dorian (2019) continue to pose challenges to track prediction. Sustained HFIP research and development may be necessary for further improvements in track prediction of these outlier events.

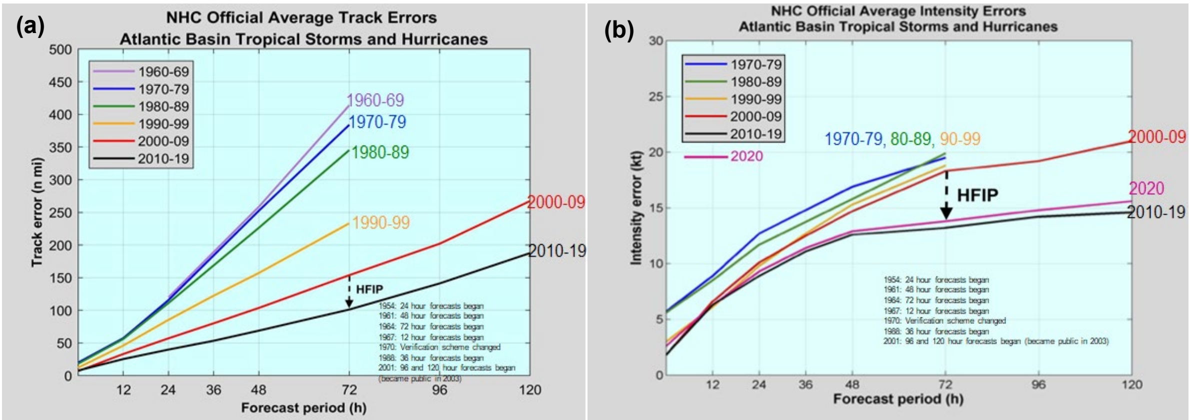

Figure 2. Official NHC (a) Track errors (1960-2019) and (b) Intensity errors (1970-2019) in the AL basin. The downward arrow denotes the period when HFIP is active.

While significant track improvements have been achieved since the 1960s, Figure 2b illustrates little or no improvement in the accuracy of NHC's official intensity forecast, until the onset of HFIP in 2009. 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 Contiguous United States (CONUS) counterparts. The HWRF and Hurricanes in a Multi-scale Ocean-coupled Non- hydrostatic (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). The Basin-Scale HWRF 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.

5. Operational HWRF and HMON systems (Stream 1)

5.1 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. A joint development between NOAA research and operations, with significant support from the Developmental Testbed Center (DTC), University Corporation for Atmospheric Research (UCAR), and the community, HWRF 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 (NMME) dynamic core and is coupled to Princeton Ocean Model (POM) and HYbrid Coordinate Ocean Model (HYCOM). It is a part of the WRF infrastructure, but using NMME dynamic core (Biswas et al., 2018; Tallapragada et. al., 2014). Improvements to model nesting, resolution (3 km in 2012, 2 km in 2015, and 1.5 km in 2018), physics, and initial conditions enhanced with aircraft observations - all coordinated under HFIP - have led to progress in improved numerical guidance.

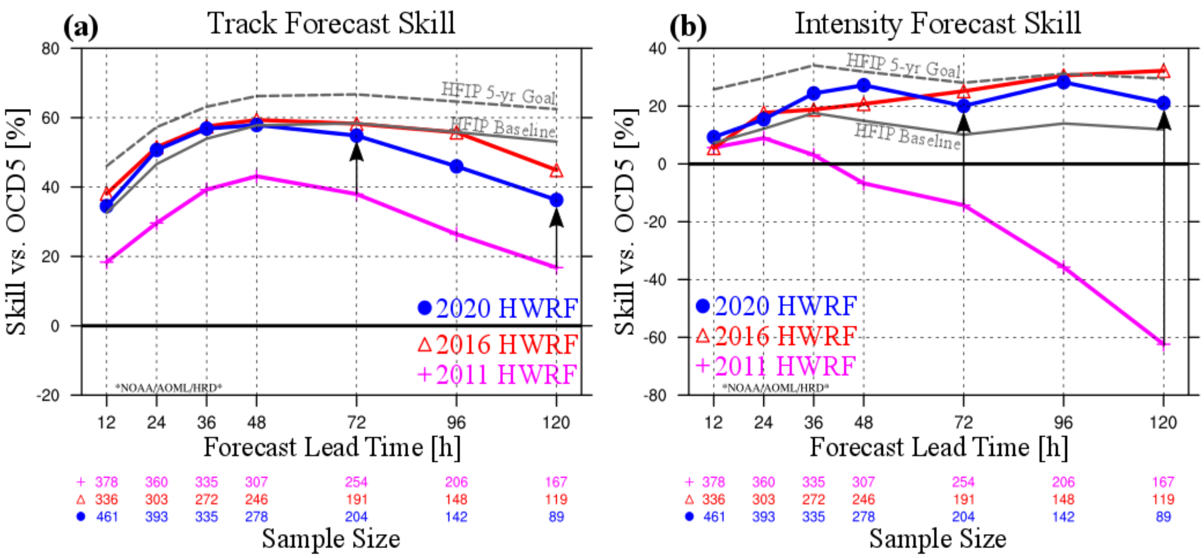

Figure 3. For the 2011 (pink cross), 2016 (red triangle), and 2020 (blue circle) North Atlantic hurricane seasons, the following is shown: (a) HWRF track forecast skill relative to NHC's climatology-persistence skill baseline (OCD5), and (b) HWRF intensity forecast skill relative to OCD5. The HFIP baselines (solid) and HFIP 5-year goals (dashed) are shown in gray. Black, vertical arrows indicate improvement of forecast skill at 72 h and 120 h lead times. 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 relative to climatology and persistence (i.e., CLIPER) from the HWRF system from 2011 to 2020. Although HWRF was initially developed to improve intensity guidance, the model also provides track guidance that is complementary to the GFS. Track performance from HWRF improved by ~20% at most of the forecast lead times between 2011 and 2020.

Figure 3b portrays the progress of HWRF in forecasting maximum intensity, measured in terms of skill relative to climatology and persistence (i.e., Decay-SHIFOR). Through 2011, HWRF operated with a single 9 km-resolution moving nest that could automatically track hurricanes[5] (Gopalakrishnan et. al., 2006). In the last nine years (2012-2020), the HWRF system was upgraded considerably under HFIP every year.

- In 2012, 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; Gopalakrishnan et. al., 2012; Gopalakrishnan et. al., 2013; Goldenberg et. al., 2015).
- In 2013, upgraded physics and vortex initialization were adopted.
- In 2014, HWRF was run in real-time in all global basins beyond the North Atlantic.
- In 2015, 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.

[5] 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.

- The year 2016 was the watermark year for 5-year improvements. New Simplified Arakawa-Schubert (SAS) and GFS-EDMF (Eddy Diffusivity Mass Flux) physics suites were implemented during this year.
- Supported by HFIP, a dramatically improved DA system was implemented in operational HWRF in 2017.
- In 2018, 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 National Weather Service (NWS).
- Due to the NCEP Central Operations (NCO) moratorium, HWRF was not operationally upgraded in 2019.
- In 2020, HWRF was upgraded to two-way ocean coupling, one-way wave-model coupling and the high-resolution land-sea masks for the moving nests.

Clearly, remarkable progress is being made under HFIP, with HWRF intensity forecasts improving by 30-80% at 48-h and longer lead times from 2011 to 2020. In fact, 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 2020 hurricane season, HWRF was the best dynamical model for the intensity forecast over the Atlantic at all lead times. HWRF performance was outstanding in the Gulf of Mexico. However, the outlier events (RI/RW) continue to impact HWRF performance from year to year, for example, Hurricane Dorain in 2019 was particularly a challenging forecast. It should be noted that forecast skill may vary from one year to the next depending on the number of cases at each lead time and the degree of difficulty that the TCs within each season present.

Despite the notable improvement to HWRF intensity predictions over the last decade, forecast skill has only met about 50% of the targeted HFIP intensity goals. Part of the reason may be associated with the lack of progress with dynamical guidance until 2012. Also, predicting RI continues to be a challenge, especially the timing of RI. Figure 4a shows the HWRF performance of RI since the 2011 season in both Atlantic and east Pacific basins. The total number and the number of RI forecasts for that year is also provided in the x-axis. It should be noted that the RI events are rare, but assume special significance especially if those occur one or two days before landfall. In order to understand the model performance in predicting these extreme TC events, we have also included the eastern North Pacific basin (Figure 4b). Although some good progress is being made in predicting RI events with the HWRF system (see the POD), false alarms continue to pose challenges to intensity predictions. Sustained HFIP research and development may be necessary for further improvements in intensity prediction of these outlier events.

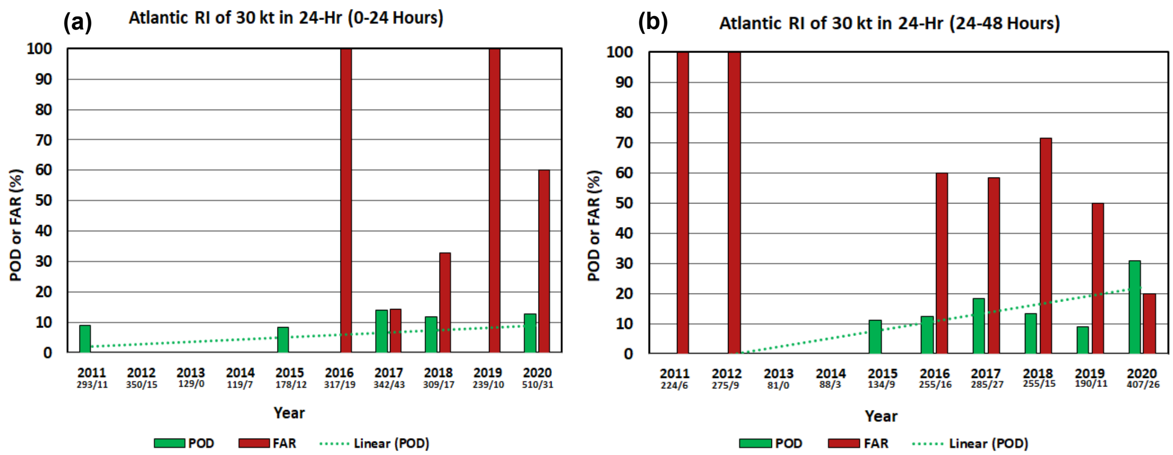

Figure 4. HWRF performance in RI predictions for 2011-2020 in the North Atlantic and eastern North Pacific basins at forecast lead times of (a) 0-24 h and (b) 24-48 h.

5.2 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 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 Common Community Physics Package (CCPP). 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.

6. 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 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 6.3).

6.1 Track Guidance

In 2020, official Atlantic track forecasts (Figure 5a) were very skillful and close to the best-performing consensus aids - FSSE (Florida State University Super-Ensemble Corrected Consensus), HCCA and TVCA (Track Variable Consensus of at least two forecasts) (Cangiolosi, 2020). GFSI (GFS with 6-hour interpolation) was the best dynamical model from 36 h to 72 h. AEMI (GEFS with 6-hour interpolation), EMXI (ECMWF with 6-hour interpolation), HMNI (HMON with 6-hour interpolation) and CTCI (COAMPS-TC 6 hour interpolation) came in second place, very close to one another. HWFI (HWRF with 6-hour interpolation), EGRI (UKMET 6-hour interpolation) and CMCI (Canadian Global Model 6-hour interpolation) were the next best models. NVGI (Navy Global Environmental Model 6-hour interpolation) lagged behind other models. It should be noted that in 2020, GFS outperformed EMXI at most of the lead times.

In the eastern Pacific (Figure 5b), the official forecasts were very skillful, close to TVCE, HCCA, FSSE, the best consensus model. HMNI, EGRI were best at 12 h and 24 h, EMXI was best at 36-60 h and 120 h, GFSI was best at 72 and 96 h. HWFI and CMCI were not very good, and NVGI lagged behind other models.

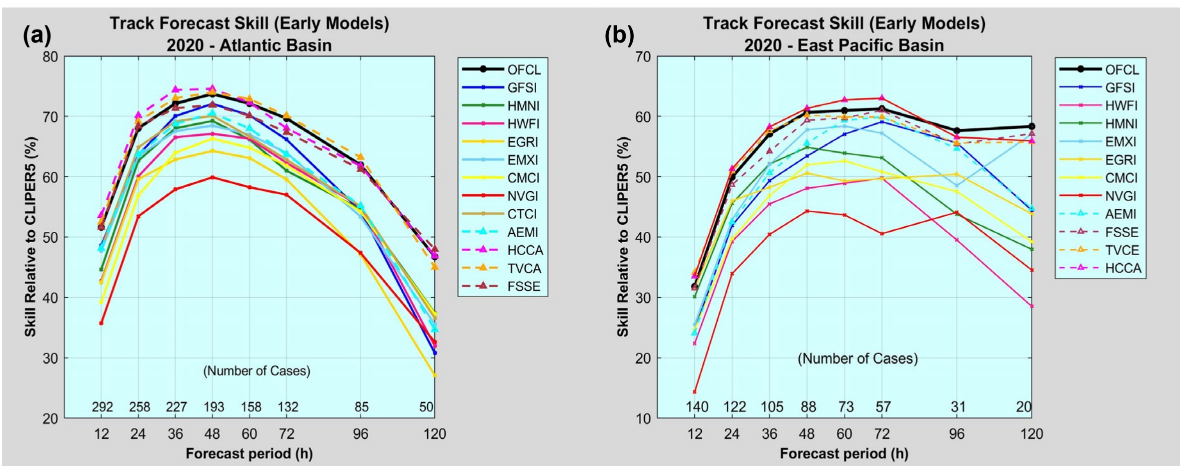

Figure 5. Official track forecast skill in 2020 for the (a) Atlantic (left) and (b) eastern Pacific (right) basins. Numbers immediately above the X-axis show the total number of cases covered by each data point.

6.2 Intensity Guidance

Intensity forecast skill for the 2020 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, HCCA was the best at all lead times while FSSE after 60 h. HWFI was a strong performer, the best individual model at most lead times. CTCI and HMNI did not do as well as HWFI. DSHP and LGEM were fair performers, but not as good as HWFI and consensus models. GFSI was somewhat competitive and EMXI was barely skillful.

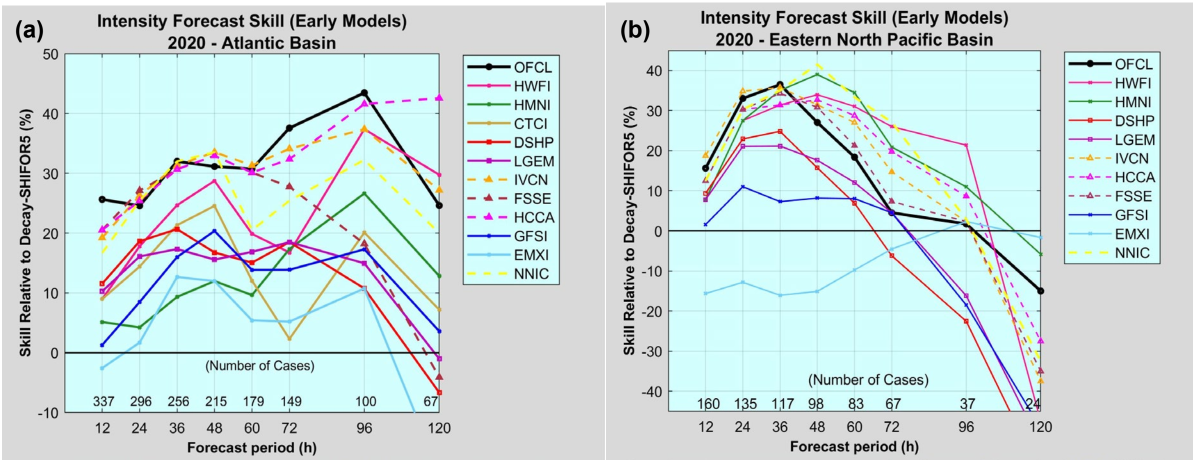

Figure 6. Official intensity forecast skill in 2020 for the (a) Atlantic Basin (left) and (b) East Pacific Basin (right). Numbers immediately above the X-axis show the total number of cases covered by each data point.

In the eastern Pacific (Figure 6b), official intensity forecast performance was good or even better than the best consensus aids (IVCN, HCCA, FSSE) through 36 h but trailed after that. HWFI was a strong performer through 96 h, but skill dropped off sharply after that. DSHP and LGEM were fair performers early but were not skillful at longer lead times. GFSI and EMXI were only competitive on days 4 and 5.

6.3 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 the subset of cases where RI was forecast or observed. The new metric is somewhat 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:

1. 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.
2. 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., HWFI, 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 is only weakly dependent on forecast lead time; the errors are high even at 24 h (26 kt) and saturate quickly. In terms of skill relative to climatology/persistence, a peak is seen from 72-96 h but skill is 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

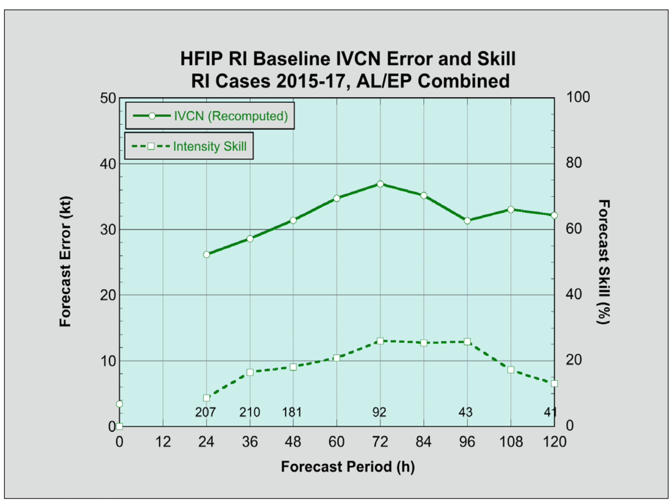

Figure 7. HFIP RI performance measures baseline errors and skill. Baseline errors are the mean absolute errors over the period 2015-17 for the Atlantic and eastern North Pacific for the variable consensus comprising at least two of the models DSHP, LGEM, GHMI, HWFI, and CTCI. Skill values are computed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

Figure 8 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 (and thus changes from year to year); in 2020 IVCN comprised DSHP, LGEM, HWFI, HMNI, and CTCI. 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 the subsequent three years 2018-20 is encouraging: at 24 h the baseline error (26.1 kt) was reduced by 27% (to 19.1 kt), at 48 h the baseline error (31.4 kt) was reduced by 19% (to 25.5 kt), and at 72 h the baseline error (36.9 kt) was reduced by 23% (to 28.5 kt).

The new RI metric does not measure rapid weakening. Methods for evaluating forecast performance for those cases will be discussed within HFIP in the coming year.

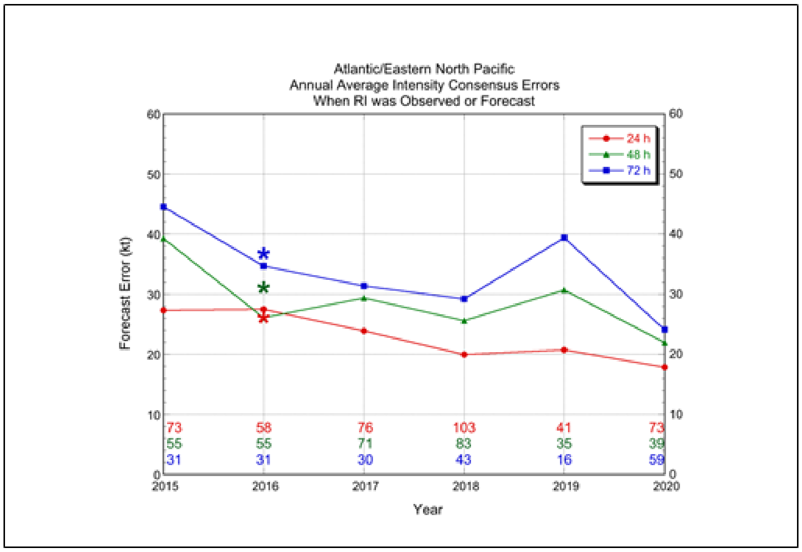

Figure 8. HFIP RI performance measure at 24, 48, and 72 h for 2015-20. The consensus evaluated for each season corresponds to NHC's operational composition of IVCN for that season. HFIP baseline errors are given by the asterisks plotted for the year 2016. Number of cases for each forecast lead are given along the bottom of the diagram.

7. Next Generation HFIP Goals and Plans

In response to Section 104 of the Weather Research Forecasting Innovation 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 Research to Operations (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.

8. 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 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 finite volume cubed-sphere (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.

HAFS comprises five major components: (a) High-resolution moving nest (b) High-resolution physics (c) Multi-scale data assimilation (d) 3D ocean coupling, and (e) Observations to support the DA.

a. High-resolution moving nest

Central to the development of HAFS is the FV3 dynamical core with an embedded moving nest capable of tracking the inner core region of the hurricane at 1-2 km resolution (cover picture). Although the FV3 model 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. Atlantic Oceanographic and Meteorology Laboratory (AOML), in partnership with GFDL and Environmental Modeling Center (EMC), is working on these developments to transition advances in HWRF to FV3-HAFS under hurricane supplemental (1A4 of the supplemental project).

b. High-resolution 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).

c. 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 developments under hurricane supplemental effort.

d. 3D Ocean coupling

The ocean model component of HAFS will use HYCOM which 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.

e. 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 such as Observing System Experiments (OSEs) and Observing System Simulation Experiments (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.

9. Important Stream-2 HREx Results

9.1 HWRF Research Advances: Multiple, Storm-following, Two-way Interactive Telescoping Nests

The operational HWRF system has shown considerable improvements under HFIP. However, the operational HWRF is a storm-centric modeling system (i.e., is configured with high-resolution nests for only one TC per forecast integration), limiting its ability to forecast storm-storm interactions, storm-environment interactions, or TC genesis. In addition, the outermost parent domain configured in the operational HWRF is small (~80° x 80°), which may inhibit further forecast skill improvements, especially beyond five days, a major goal of next-generation numerical weather prediction efforts. HWRF's configuration is incompatible with current data assimilation software because the outermost parent domain moves from one forecast to the next, posing many challenges for producing a large-scale analysis. These points may represent impediments to further advances in hurricane forecast guidance from storm-centric dynamical models like HWRF.

For this reason, an experimental version of HWRF (HWRF-B), known as the "basin-scale" or "multi-storm" configuration, was created under HFIP with some advanced configuration options: 1) a large, static outermost domain that covers approximately one-fourth of the globe, and 2) multiple sets of movable, multi-level nests, each following a different storm at a horizontal resolution on par with that in the current operational HWRF system. As a result, HWRF-B has the ability to produce high-resolution forecasts for multiple TC within the same model integration and, also, serves as a prototype for the ongoing development of moving nests within HAFS. HWRF-B allows for advanced data assimilation evaluations given its static outermost domain, and has already shown promise in OSEs and OSSEs. The outermost domain was moved north by 10° and west by 10° to accommodate more TCs in areas of responsibility for the NHC and the Central Pacific Hurricane Center (CPHC). The outermost domain is shown in Figure 9 with moving nests outlined in black. HWRF-B has been a collaborative effort between AOML and NCEP.

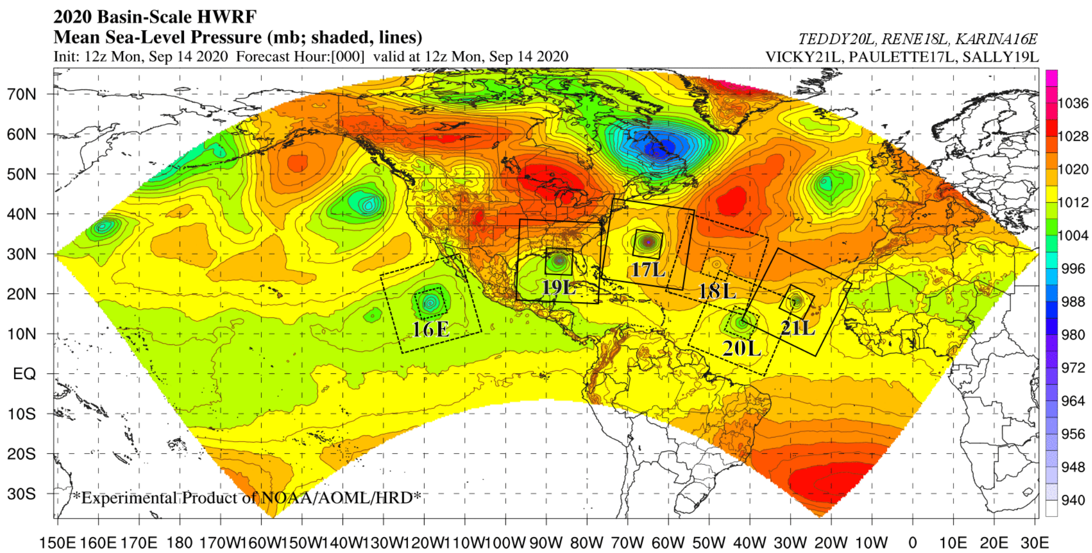

Figure 9. Mean sea-level pressure (hPa) is shown for the outermost domain (D01) from an HWRF-B forecast initialized at 1200 UTC 14 September 2020. HWRF-B was configured with high-resolution, storm-following multi-level nests (solid black boxes) for three tropical cyclones (TCs): Hurricane Paulette (17L), Hurricane Sally (19L), and Tropical Storm Vicky (21L). Three additional TCs were also active at the time shown, but were not configured with nests due to limitations in computing resources (dashed black boxes): Tropical Depression Rene (18L), Tropical Storm Teddy (20L), and Hurricane Karina (16E).

HWRF-B was an HFIP real-time demonstration for the eighth year during the 2020 North Atlantic and eastern North Pacific hurricane seasons, in parallel with operational hurricane models and HAFS experiments. A multi-storm coupler was developed, tested, and implemented in HWRF-B to exchange information between the atmosphere and ocean models for multiple storms at high resolution. HWRF-B tracked up to three TCs per forecast at high resolution (~1.5 km) during the 2020 real-time demonstration; off-season retrospective forecasts are tracking up to five TCs with high-resolution moving nests and more nests are possible with expected increases to high performance computer resources. In 2020, HWRF-B assimilated the exact same satellite, ground-based, and aircraft observations as the operational system. Off-season retrospective forecasts are configured with a self-cycled data assimilation software that is identical to the one used for the operational system. This data assimilation system was expanded to cycle multiple TCs, the first of its kind. With high cyclone activity in 2020, especially in the North Atlantic basin, HWRF-B produced forecasts for multiple TCs in > 90% of its simulations. For example, high-resolution nests in HWRF-B tracked Hurricane Paulette (17L), Hurricane Sally (19L), and Tropical Storm Vicky (21L) for a forecast initialized at 1200 UTC 14 September 2020 (Figure 9).

HWRF-B continues to show some intensity improvements over the operational HWRF, with 5-10% skill over HWRF at 72-96-h forecast lead times (Figure 11b). HWRF-B also continued to outperform the operational HWRF when multiple TCs were active in the same forecast. Storm size forecasts produced by HWRF-B showed notable improvements over HWRF, especially for 34-kt wind radii and 64-kt wind radii (Figure 11). Hurricane-force wind radii errors were less than 12 n mi at lead times of 36 h and longer (Figure 11b). Track forecast skill versus HWRF regressed slightly because the new position of the outermost domain (shifted north and west) led to unexpected biases in mid-latitude systems over North America (Figure 10a). Further analysis demonstrated that capturing fine-scale details of as many TCs as possible via multiple moving telescopic nests is important to predict realistic storm-storm interactions and, thus, to produce accurate forecasts of maximum intensity, storm structure, and track. More importantly, moving nest technology showcased in HWRF-B will be foundational to the HAFS, the next-generation hurricane application within UFS.

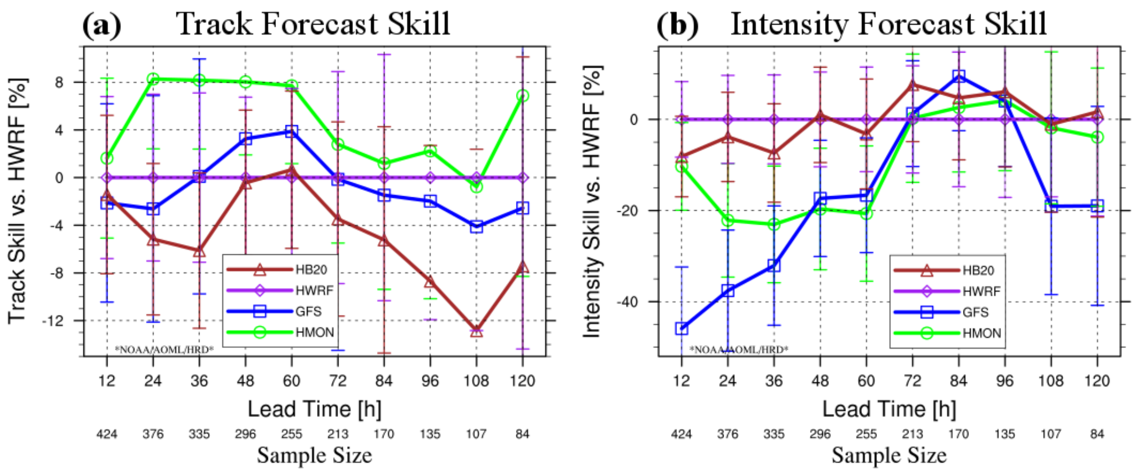

Figure 10. Verification of a) track forecast skill and b) intensity forecast skill using 2020 operational HWRF forecasts (HRWF; purple) as a baseline for the following models: 2020 HWRF-B forecasts (HB20; brown), GFS forecasts (GFS; blue), and HMON forecasts (HMON; green). The sample size at each lead time is shown at the bottom.

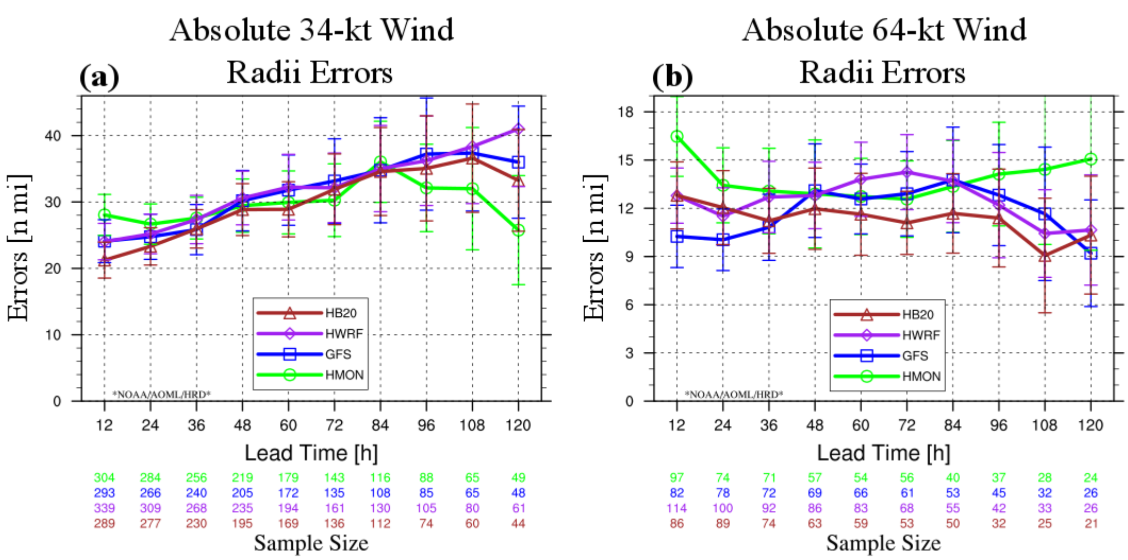

Figure 11. Verification of a) absolute 34-kt wind radii errors and b) absolute 64-kt wind radii errors using 2020 operational HWRF forecasts (HRWF; purple) as a baseline for the following models: 2020 HWRF-B forecasts (HB20; brown), GFS forecasts (GFS; blue), and HMON forecasts (HMON; green). The sample size at each lead time for each model is shown at the bottom.

9.2 HAFS Experimental systems

Four configurations of the HAFS model were run in 2020 HREx. The 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. There were two other configurations of the HAFS model tested as a part of 2020 HREx. These configurations are: uniform grid projection on Extended Schmidt Gnomonic (ESG) grid (HAFS v0.1J); and HAFS ensembles with 18 members (HAFS v0.1E). In addition, the GFDL ran the T-SHiELD (Harris et al. 2020, which uses a 2-way nested configuration similar to HAFS-globalnest) and shared their results with HFIP to provide critical inputs on the FV3 dynamic core behavior that may be useful for further advancements of HAFS.

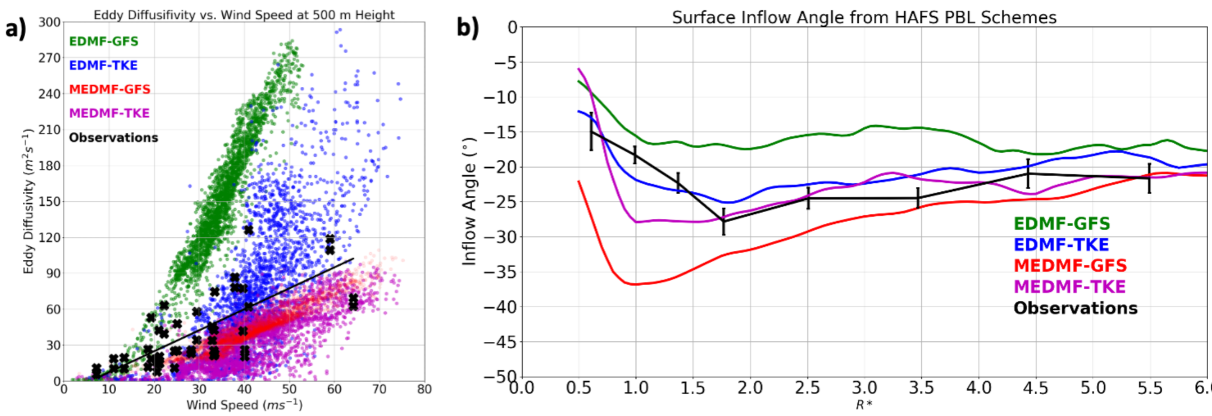

Figure 12. a) Eddy diffusivity in HAFS-globalnest from 4 different schemes and observations: Hybrid-EDMF (green), Hybrid-EDMF with α = 0.25 (red), EDMF-TKE (blue), EDMF-TKE with mixing length capped at 100 m (magenta), and observations (black). b) As in a), but for inflow angle in HAFS.

The HAFS-globalnest HAFSv0.1B experiment was designed to analyze the influence of the high-resolution nest on the global model. The HAFSv0.1B was also used to test the influence of PBL changes on the forecast. HAFSv0.1B used a modified version of the EDMF-TKE planetary boundary layer (PBL) scheme (Gopalakrishnan et al. 2021). The modifications were made based on comparison of observed eddy diffusivity and mixing length with the values in the default scheme, and involved imposing a cap of 100 meters on two mixing length scales in the scheme. Figure 12 shows how the modifications changed the eddy diffusivity and inflow angle in a simulation of Hurricane Michael (2018). For 2020, HAFS-globalnest also used 75 vertical levels, compared to the 64 in 2019.

Table 2. Model configurations for the 2020 HREx HAFS-SAR (HAFA) and HAFS-globalnest (HAFB) experiments.

Model	**Domain**	**BCs**	**Finest Grid Spacing**	**Vertical Levels**	**PBL**	**Micro-physics**	**Convective Scheme**	**Ocean Model (Y/N)**
HAFS-Global nest (*HAFB*)	Global With Atlantic Static Nest	Global HAFS (2-way)	3 km	75	Modified EDMF-TKE	GFDL	Global: Scale-aware SAS Nest: None	None

HAFS-LAM (*HAFA*)	Atlantic Static Nest	GFS	3 km	91	Modified EDMF	GFDL	None	Regional HYCOM

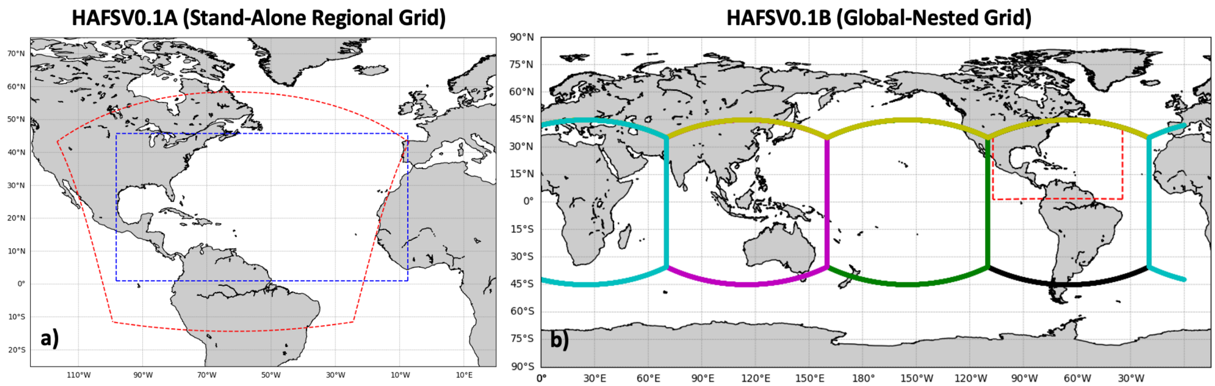

Figure 13. a) Regional 3-km FV3/HAFS-A atmospheric domain (red box) with ocean HYCOM domain (blue box). b) Global FV3/HAFS-B with embedded 3-km static-nest (red box).

For the first time the regional version of HAFS (HAFSv0.1A) was run with active ocean coupling during the 2020 season (Figure 13a). HAFS-SAR used a regional version of HYCOM (Bleck 2002). To maintain computational efficiency, a slightly smaller Atlantic nest was used for HAFS-globalnest in 2020, which made the domain of HAFS-globalnest slightly smaller than HAFS-SAR (Figure 13b). Also there were other differences in configuration between the global and regional versions of HAFS. Instead of the modified EDMF-TKE scheme, it still used the Hybrid EDMF scheme, similar to what is in HWRF and HAFSv0.1A. The convection scheme was turned off in the 2020 real time HAFS-A experiment. In addition, HAFS-SAR used more vertical levels for the 2020 season, with a 91-level configuration that featured more levels in the PBL but was slightly different from that used in HAFS-globalnest. Table 1 highlights the configurations of HAFS-globalnest and HAFS-SAR.

Figure 14a,c shows the track forecast errors and skill from HAFS-globalnest, HAFS-SAR, the GFDL T-SHiELD (Harris et al. 2020, which uses a 2-way nested configuration similar to HAFS-globalnest), and three operational models (GFS, HWRF, HMON). HAFS-globalnest had the lowest track errors for most forecast hours and notable skill relative to HWRF**,** while HAFS-SAR was less skilled, especially at later leads. HAFS-globalnest has the highest intensity errors, especially at longer lead times (Figure 14b,d), driven mostly by a positive bias. Ongoing evaluation of the PBL scheme and implementation of a DA system in HAFS is expected to improve analysis and prediction of TC structure and intensity.

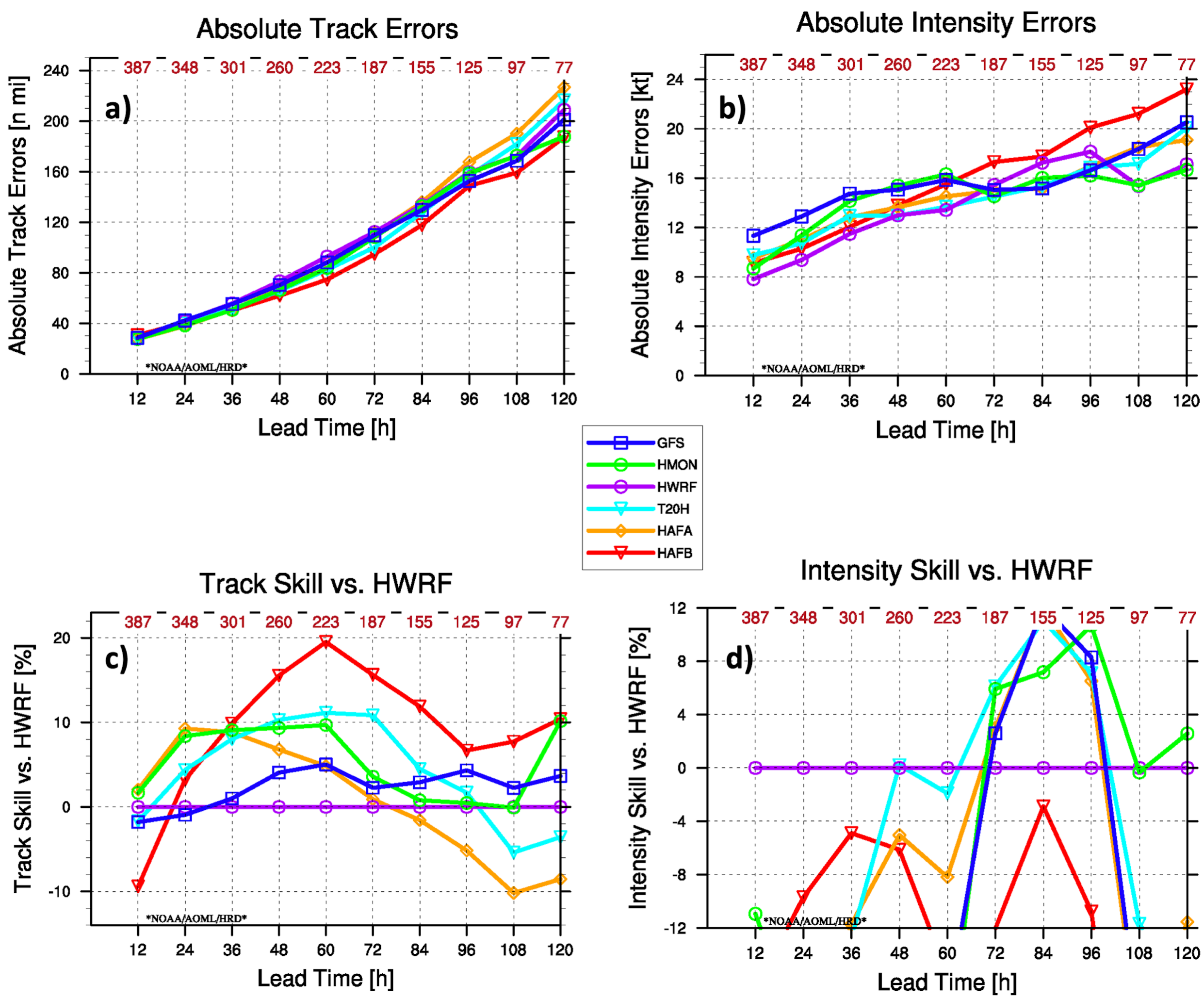

Figure 14. a) Track forecast errors (n mi) from 0-120h for HAFS-globalnest (red), HAFS-SAR (orange), GFDL T-SHiELD (light blue), the operational HWRF (purple), the operational HMON (green), and the operational GFS (blue). b) As in a), but for intensity errors (kt). c) Track forecast skill for all models relative to operational HWRF. d) As in c), but for intensity skill.

An evaluation was performed for the first time during the 2020 season to demonstrate how the feedback from the 3-km nest in HAFS-globalnest impacts the overall global skill of the global-nested model. Figure 15 shows the 500-hPa global anomaly correlation for the period covered during the 2020 hurricane season for the global GFS and HAFS-globalnest. The results are nearly identical, indicating that while the high-resolution nest does not seem to significantly improve the global skill, it also does not seem to be causing any significant degradation either. Work is ongoing to evaluate and improve regional biases that are noted, including a weak bias in the subtropical ridge over the Western Atlantic.

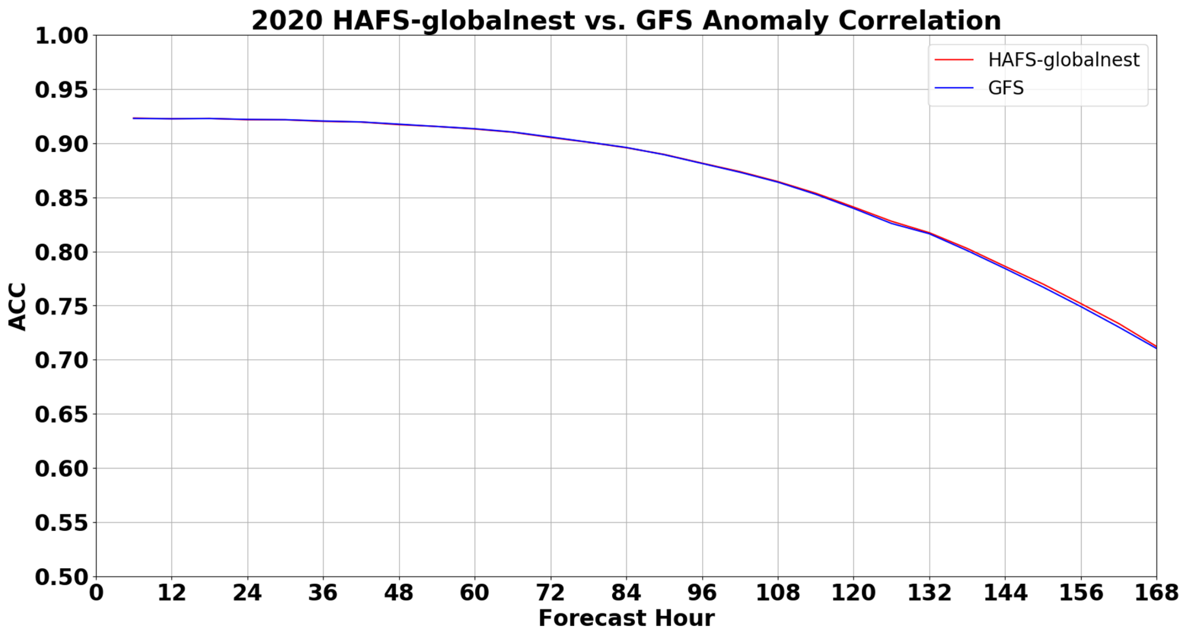

Figure 15: Global 500-hPa height anomaly correlation from HAFS-globalnest (red) and GFS (blue) for the 2020 season.

Comparing all HREx configurations with operational HWRF (Figure 16), all four HAFS configurations did well with track skill in the Atlantic basin, while HAFS v0.1B had the best track skill. HAFS v0.E track skill was comparable with both HAFS v0.1A and HAFS v0.1J, until day 3. As for the intensity skill, all HAFS versions lagged behind HWRF. Of all HAFS versions, HAFS v0.1E had the lowest intensity errors for day 3-5. In Eastern Pacific basin, HAFS v0.1A and HAFS v0.1J had the best track skills. For the intensity skill, again all HAFS versions lagged behind HWRF.

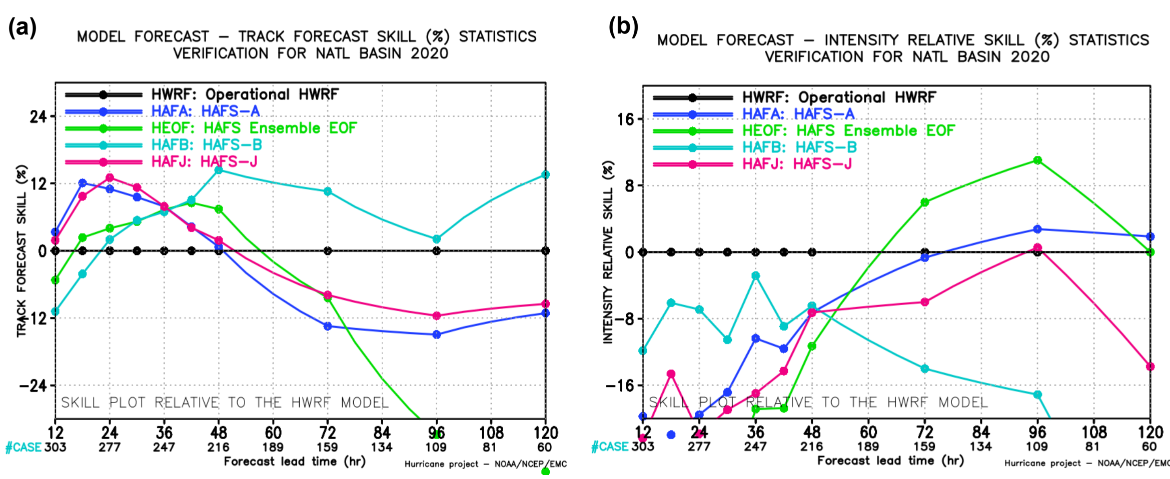

Figure 16. (a) Comparison of track forecast skills; and (b) intensity forecast skills, for regional HAFS (navy blue), global HAFS (cyan), HAFS ESG grid (red) and HAFS ensembles (green) vs. HWRF (black) from 0-120h.

10. New Products, Tools, and Services at NHC

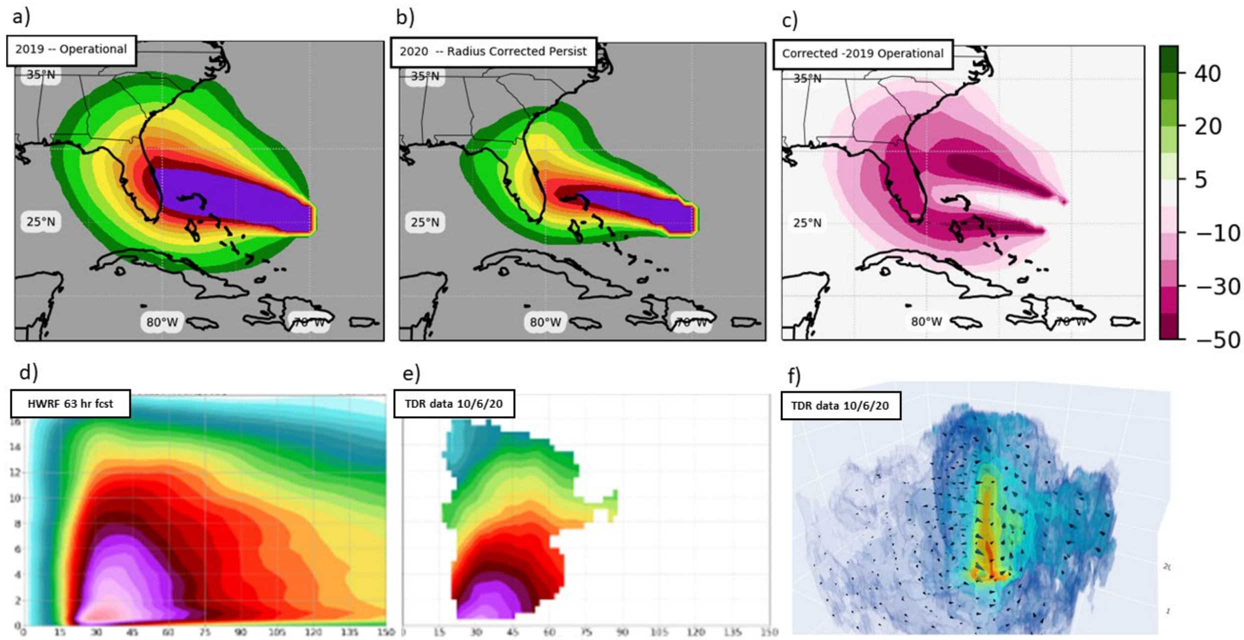

Figure 17: Examples of PPAV team products and results from 2020: a) Operational wind speed probability model output for Hurricane Dorian (2019) b) Experimental Radius-Corrected wind speed probability model output for the same forecast c) Difference plot between the operational and experimental models. The experimental probabilities were lower and more realistic at locations relatively far from the center of Dorian. d) Sample HWRF forecast of Hurricane Delta (2020) showing axisymmetric vertical structure. e) Tail Doppler observations of Delta valid at approximately the same time. F) 3-D image of Tail Doppler observations of Delta at the same time.

10.1 Operational and Real-Time Applications

A number of enhancements to operational tools and applications were developed with HFIP support in 2020. One of the most significant was an upgrade to the P-Surge model. The upgrade improved the representation of storm size and forecast uncertainty in the ensemble. Work began toward another upgrade, which was implemented in Spring 2021, to improve the ensemble representation of the radius of maximum winds, an important parameter for storm surge prediction. NHC and CIRA tested new methods for statistical post-processing of dynamical model forecasts, including machine learning techniques, to improve intensity prediction. Preliminary verification from 2020 indicated that the new methods have skill at forecasting intensity.

Several projects also focused on improvements to public forecast products and warnings from NHC and NWS. NHC and CIRA developed a method to improve the representation of storm size to better match official forecasts in the wind speed probability model (Figure 17 a-c). This method was evaluated in 2020 and shows promise based on results from past tropical cyclones. Another effort to convert wind speed probability output to tropical storm and hurricane watch/warning guidance for NWS forecasters is underway, and several tests were conducted in 2020. A Hazardous Weather Testbed experiment was conducted with several NWS Weather Forecast Offices (WFOs) in Spring 2021 to demonstrate the watch/warning guidance in AWIPS. Another project to develop products to express uncertainty in the intensity forecast is also underway. Several HFIP-developed products such as the time of arrival graphic and the “Be Ready By” graphic continued to be supported in 2020.

HFIP also supported improvements to operational infrastructure. The ATCF was updated in 2020 to meet new NHC operating system needs, add 60-h forecast track, intensity, and 34-kt and 50-kt wind radii information to NHC's public products, improve depictions of forecast confidence for the forecasters at day 6 and 7, enhance backup capabilities, and improve the ingest of "fix" data from hurricane hunter aircraft. Experimental HAFS and basin-scale HWRF forecasts were available in the ATCF for the first time in 2020 and other experimental products were made available to NHC forecasts through a new experimental portal at NHC.

10.2 Display and Diagnostic Activities

The HFIP community continued to improve model diagnostics and visualization techniques in 2020. HRD improved methods for visualizing tropical cyclone vertical and horizontal structure, and directly compared model output to Tail Doppler Radar (TDR) output (Figure 17 d,e). HRD also developed and maintained a website for viewing real-time output from the experimental basin-scale HWRF. Finally, HRD continued to develop methods for displaying and evaluating planetary boundary layer fields from models and comparing them with fields from dropsondes. Other web and browser-based visualization tools from ESRL and NCAR were maintained and improved, including real-time products on hfip.org. A browser-based diagnostic tool developed by NCAR was upgraded at NHC and available to NHC staff in real time, both on site and remotely. NCAR also maintains a public version of the tool, available through the HFIP website. The public version of the NCAR tool has been used for training meteorologists at the World Meteorological Organization Region IV workshop on tropical cyclones. An effort is underway by HFIP to re-unify many of the web-based visualization tools.

10.3 Experimental Projects

The hurricane supplemental enabled some accelerated initial developments of HAFS. One project is an effort to examine the calibration of the Stepped Frequency Microwave Radiometer (SFMR) instrument at very high wind speeds. Datasets required to do the calibration have been updated and improved, and work is underway to identify a potential bias at very high wind speeds. Another project is focused on the use of 3-D visualization software for model and other tropical cyclone data (Figure 17 f). Real-time 3-D imagery from the NOAA P-3 aircraft were available to the forecasters in 2020 and was used operationally to diagnose vertical tilt during the early intensification stages of Hurricane Laura. The third project extended NHC's statistical forecast guidance to 7 days and modernized the code to help with long-term maintenance and future development. 7-day DSHP, LGEM, and SHIP forecasts were available to NHC forecasters in 2020, including the associated diagnostic files.

10.4 Community Involvement

R2O was one of the initial goals of the Weather Research and Forecasting (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[6]. 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.

<u>HFIP Ensemble Diagnostic Products</u>: NCAR continues to focus on new diagnostic products to help evaluate the performance of multi-ensemble hurricane track and intensity forecasts. During this past year, NCAR has explored several new visualizations tools, which show forecast tracks stratified by a selected diagnostic field. The stratified tracks are colored by the diagnostic parameter. Possible diagnostic parameters could include environmental conditions (e.g., storm intensity, environmental vertical wind shear, maximum potential intensity, sea surface temperature, etc.) or inner core storm structural characteristics (precipitation symmetry, radius of maximum winds, inertial stability, etc.). The diagnostic ensemble evaluation tools use parameters available from the ATCF a-decks, such as intensity, minimum sea level pressure, and forecast lead-time. Figure 18a shows an example visualization for Tropical Storm (TS) Sally. The diagnostic parameter is the ensemble member predicted storm intensity (MSLP). The analysis shows that cases where the storms going to the right of track were more intense, and all of the tracks going further west were weaker for TS Sally. Figure 18b shows another example of a diagnostic evaluation. In this case, the forecast tracks are color-coded by lead-time. The ensemble members provide a visual evaluation of the track variability, which can help evaluate the track forecast uncertainty for different lead-times.

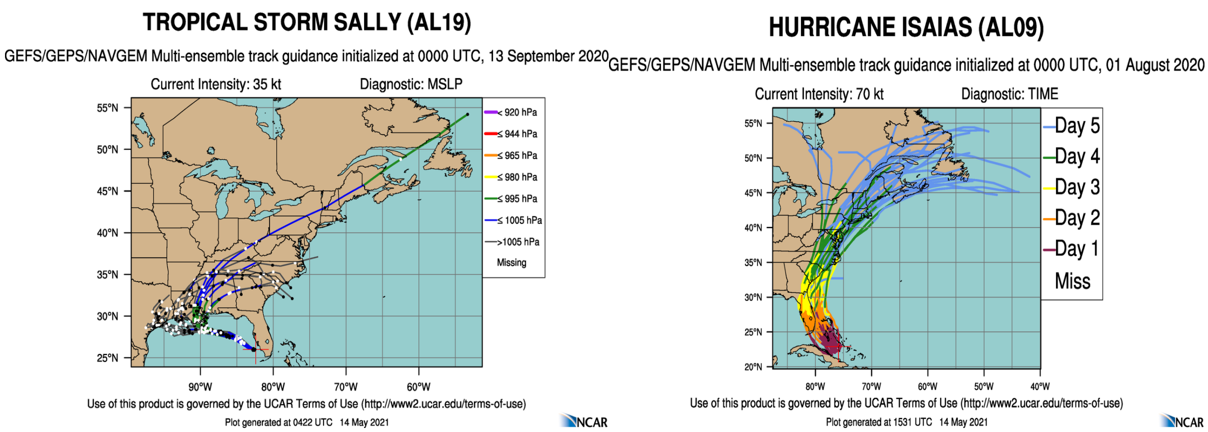

Figure 18. (a) Intensity diagnostic plot for the late-cycle forecast track guidance for TS Sally, initialized at 00 UTC on 13 September 2020. (b) Lead time diagnostic plot for the late-cycle forecast track guidance for Hurricane Isaias, initialized at 00 UTC on 01 August 2020.

<u>NCAR NHC Display System:</u> The hurricane display and diagnostic capabilities allow operational forecasters and research scientists to explore the performance of hurricane track and intensity forecast models for current and historical cases. The system has been built upon modern and flexible technology, including OpenLayers Mapping tools and an efficient MySQL database. During the past year, the development has focused on improving the capabilities and functionality of the display system and diagnostic system. An example of the updated system is shown in Figure 19. The plot shows the diagnostic evaluation of forecast tracks stratified by forecasted storm intensity for Hurricane Dorian. In addition to gridded sea-surface temperature fields. The system also displays GFS derived gridded fields including wind shear, moisture, and precipitable water.

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

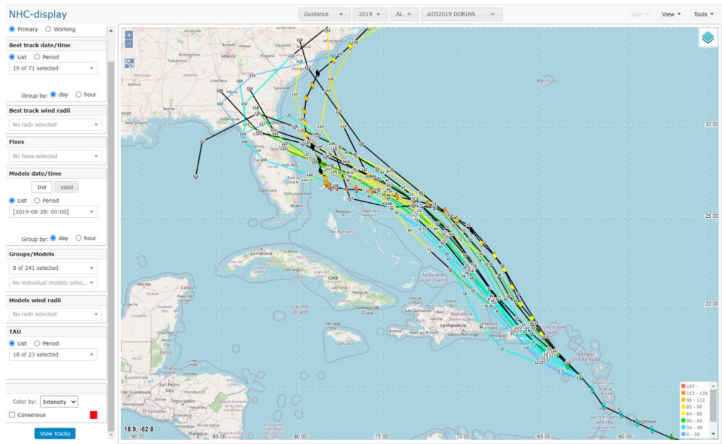

Figure 19. Lead time diagnostic plot for the late-cycle forecast track guidance for Hurricane Dorian, initialized at 00 UTC on 28 August 2019.

11. NOAA Federally Funded Opportunity (FFO)

The following tables provide the list of projects supported by HFIP during 2018-2020 and 2020-2022.

Table 3. HFIP Supported Projects from 2018-2020.

HFIP Collaborative Awards Round V (2018-2020)		
PI Name	**PI Institution**	**Project Title**
Agnes Lim	University of Wisconsin (UWI)	Advanced DA Techniques for Satellite-Derived Atmospheric Motion Vectors from GOES 16/17 in the HWRF
Andrea Schumacher	Colorado State University (CSU)	Using Dynamically-Based Probabilistic Forecast Systems to Improve the NHC Wind Speed Products
Kerry Emanuel	Massachusetts Institute of Technology (MIT)	New Frameworks for Predicting Extreme Rapid Intensification
Ping Zhu	Florida International University (FIU)	Rapid Intensification Changes: Improving Sub-Grid Scale Model Parameterization and Microphysical-Dynamical Interaction
Ryan Torn	SUNY Albany	Evaluating Initial Condition Perturbation Methods in the HWRF Ensemble Prediction System
Ting-Chi Wu	Colorado State University (CSU)	Enabling Cloud Condensate Cycling for All-Sky Radiance Assimilation in HWRF

Table 4. HFIP Supported Projects from 2020-2022.

HFIP Collaborative Awards Round VI (2020-2022)		
PI Name	**PI Institution**	**Project Title**
Alan Brammer	CSU-CIRA	Extending the Tropical Cyclone Genesis Index to Global Ensemble Forecasts
Enrique Curchitser	Rutgers	Developing Regional Ocean Modeling Capabilities with MOM6 for use in the UFS
Ryan Torn	SUNY Albany	Application of Innovation Statistics to Diagnose Biases in the HAFS System

12. Socio-economic Aspects of HFIP

NHC's tropical cyclone forecast track graphic, commonly referred to as the cone of uncertainty (referred to as the "cone"), may be both the most viewed and most misinterpreted product within the tropical cyclone product suite. Designed to convey the forecast uncertainty of the center of a tropical cyclone's track, the cone's visual features have come under scrutiny with many studies and reports pointing to misunderstanding. The NOAA Hurricane Charley Service Assessment (2006) documented how residents and emergency managers focused too much on the original skinny black line, discounting the geographic areas in the surrounding cone as not at risk to the hurricane's associated hazards. The NHC later set the default version of the graphic to exclude the skinny black line allowing users to toggle that feature on/off if they choose. However, the issue of misinterpreting the line, or one's mental interpolation of a line between forecast points, persists as noted in the more recent NOAA Hurricane Matthew Service Assessment (2017). Beyond the skinny black line, many users also anchor to whether they are "inside" or "outside" of the cone to make decisions. Since the associated hazards of a tropical cyclone often extend well beyond the bounds of the cone, the use of the cone in this way is disconcerting and potentially dangerous.

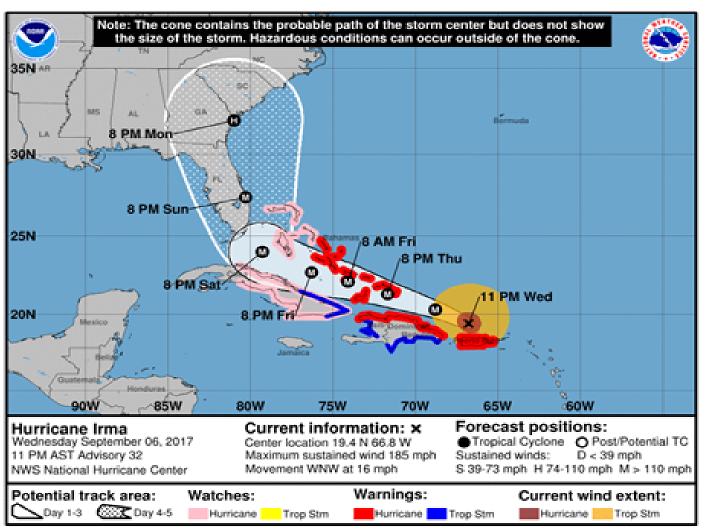

Figure 20. The 5-day cone of uncertainty with black track line toggled off.

Misinterpretation may exist because the cone of uncertainty conveys a lot of complex hurricane information. The cone shape represents an outline of the 67th percentile of NHC's average track effort over the last 5 years at each forecast time. This means that the size of the cone is not dynamic on a storm-by-storm basis, or even a forecast-by-forecast basis, and reflects the amount of error (forecast vs. actual path) averaged for all events over the previous five years. An implicit function of the cone is to give people a "heads up" that they may need to prepare for a tropical cyclone based on their proximity to the shaded area, but the cone does not convey the specifics of each hazard associated with the tropical cyclone.

Despite the cone's limitations, it remains one of the most popular public-facing NHC products. Broadcast meteorologists and the private weather industry often make their own version of the cone of uncertainty, showing it on television as well as posting it online. The appeal of the cone is that it helps answer the question, "Where is the hurricane going?," providing a succinct visual summary of the storm's likely forecast track and intensity. In some regards, it is the "go-to" product for many users.

Because of these long-term misunderstandings and the importance of conveying risk and uncertainty, NWS commissioned a study in 2018 to focus on the cone of uncertainty and the related information it conveys. The research specifically examined - i) How do people interpret (or misinterpret) the 'Cone Graphic'? ii) How integral is the cone graphic to international partners' decision-making? iii) How much do important economic sectors rely on the 'Cone Graphic' for operational decision-making? iv) Does the Cone Graphic meet these users' and stakeholders' needs?

The literature review examined more than 50 studies and reports to glean insights into interpretations, uses, and perceived strengths and weaknesses of the 'Cone Graphic' by members of the public, public officials, and broadcasters. The interviews gathered information about how international meteorologists use and communicate the 'Cone Graphic', along with their suggestions for enhancements. The survey provided important insights into the interpretation and use of the 'Cone Graphic' by 491 representatives of four key industry sectors such as transportation, marine, tourism/recreation, and energy/utilities, that can incur high economic impacts and losses due to tropical cyclones.

The key findings across all research are:

- The 'Cone Graphic' is very familiar to all audiences, from members of the public to experienced professionals in the industry sectors studied.
- The graphic is valuable (even critical) to decision-making for international forecasters and industry sector professionals.
- Aspects of the cone graphic are misunderstood or misinterpreted, even by experienced professionals (especially those in the tourism/recreation sector and those without prior storm experience).
- The 'Cone Graphic' is not the only piece of information that people (members of the public and professionals alike) use to make preparedness decisions.
- There may be some over-reliance on the 'Cone Graphic' for some groups, including energy/utility entities and those lacking prior storm experience.
- Hurricane graphics may not play a key role in decision-making for members of the American public (not necessarily true for some international audiences, however).
- The 'Cone Graphic' might have too much information to process cognitively—and yet it does not depict some of the parameters that people want most to inform decision-making, such as storm size, intensity, confidence, where the worst weather could be, what the potential impacts might be, what the calls to action are, and what hazards are of most concern.

There are several paths forward that the NHC can consider, such as:

i. Increase education around the current ‘Cone Graphic’ and assess the success of these measures.
ii. Make cosmetic improvements to the current visualization (e.g., modernize the look, reduce the graphical elements) and/or messaging around the graphic.
iii. Use technology and data visualization tools to enhance the utility and user-appeal of the graphic.
iv. Consider if alternative visualizations could better convey uncertainty information.
v. Consider the possibility of developing supplemental and/or combination graphics depicting the type of information people want to see to help them make preparedness decisions.

Looking ahead, the NHC may want to consider if there are ways to address the needs of sophisticated and unsophisticated users alike, to test potential changes or prototypes with user groups, and to consider the best approaches for implementing any changes it may be considering given how familiar the graphic is to so many people.

13. 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 address the intensity forecast problem because the horizontal resolution in global models are incapable of capturing the hurricane eye wall 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 40-60% 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.

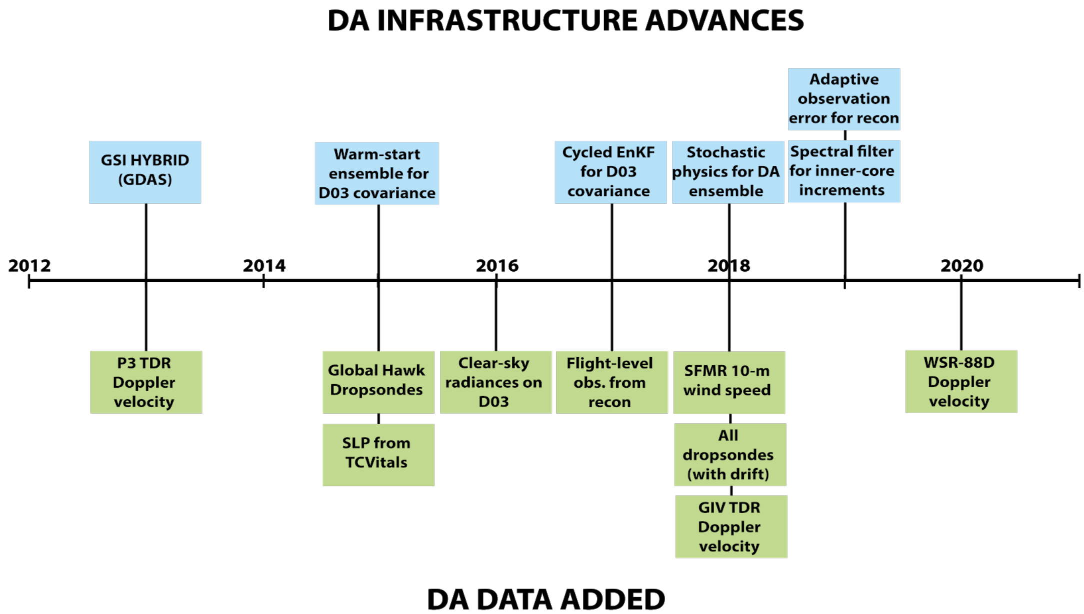

Figure 21. Evolution of inner-core data assimilation techniques under HFIP.

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 21 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 (Figure 9). The Ocean-Coupled Basin-Scale HWRF, which was run in Stream 2 for the past 3-4 seasons, is starting to demonstrate how basin wide domain with multiple-moving nests tracking several storms simultaneously in AL and EP basins could improve storm-storm and land-storm interactions without using uniform high-resolution domain, hence providing an operational solution for the TC forecasting (Figure 9). 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 provide another step in improvements to 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, originated as Stream 2 activity, and supported in a real-time demonstration mode during the hurricane season and then transitioned to operations. Beside these, 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 2020, we are only half way through in terms of improvements.

7 http://hfip.org/documents

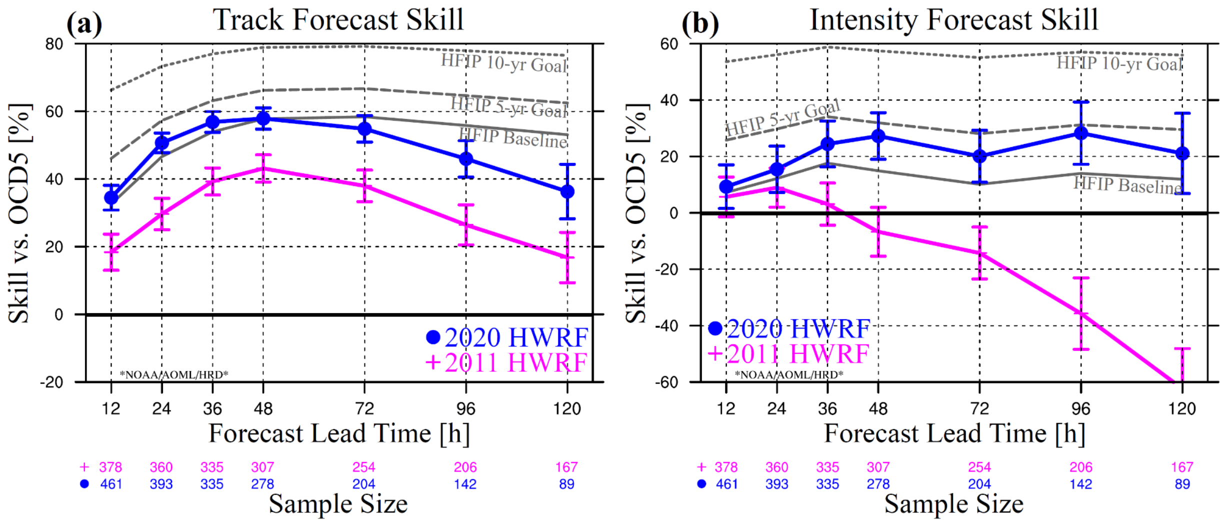

Figure 22. (a) 2011 HWRF (pink cross) and 2020 HWRF (blue circle) track forecast skill relative to NHC's climatology-persistence skill baseline (OCD5) for the North Atlantic basin, and (b) as in (a), except for intensity forecast skill relative to OCD5. The HFIP baselines (solid), HFIP 5-year goals (long dash), and HFIP 10-year goals (short dash) are shown in gray. Sample sizes for the two versions of HWRF are provided beneath each panel for verified forecast lead times.

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 15-20% improvement of track forecast skill (Figure 22a), and more than 40-60% improvement of intensity forecast skill (Figure 22b) 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 22b). Yet, as shown in Figure 22b, these improvements in intensity predictions only resulted in reaching 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 22b) 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 the original HFIP baseline (Figure 22a). It appears that global models with two-way interactive high-resolution nests may be the ultimate solution for both track and intensity predictions (Figure 14a and c). Moreover, our needs for additional forecast improvements and products have grown since 2009.

The Hurricane Forecast Improvement Project, authorized by The Weather Research and Forecasting Innovation Act of 2017, aims to further improve hurricane forecast accuracy, lead time, and risk communication required to save lives, minimize damage, and protect the livelihoods of vulnerable populations.

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).

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 in HREx for the 2020 hurricane season. The configurations are (i) the high resolution ocean coupled regional model (HAFS v0.1A); (ii) global model with a high resolution nest (HAFS v0.1B); (iii) uniform grid projection on Extended Schmidt Gnomonic (ESG) grid (HAFS v0.1J); and (iv) HAFS ensembles with 18 members (HAFS v0.1E). All HAFS configurations did well with track skills in the Atlantic basin while HAFS v0.1B had exceptional track skills (Figure 14a). As for the intensity skills (Figure 14b), all HAFS versions lagged behind HWRF illustrating the need for further developments of physics and inner core DA techniques consistent with HWRF.

14. 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 funds to improve weather forecasting, hurricane intensity forecasting and flood forecasting and mitigation capabilities to support HAFS developments under HFIP from 2019-2022. This provides a firm start 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.

References

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., (2020): National hurricane center 2020 Preliminary forecast verification. The 12th HFIP Annual Rev. Conf. presentation, Nov 17, 2020, Virtual. [Available online at http://hfip.org/sites/default/files/events/65/110pm-cangialosi-nhc-verification.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.

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.

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

Hazelton A. T. 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.

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: 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 Centre 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
DSHP	Decay Statistical Hurricane Intensity Prediction Scheme (SHIPS) model
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
HREx	Hurricane Real-time Experiment
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
OSTI	Office of Science and Technology Integration
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
UCAR	University Corporation for Atmospheric Research
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
UW-NMS	University of Wisconsin Non-hydrostatic Modeling System
WCOSS	Weather & Climate Operational Supercomputing System
WMO	World Meteorological Organization
WRF	Weather Research & Forecasting
WFO	Weather Forecast Office
WPO	Weather Program Office