NOAA

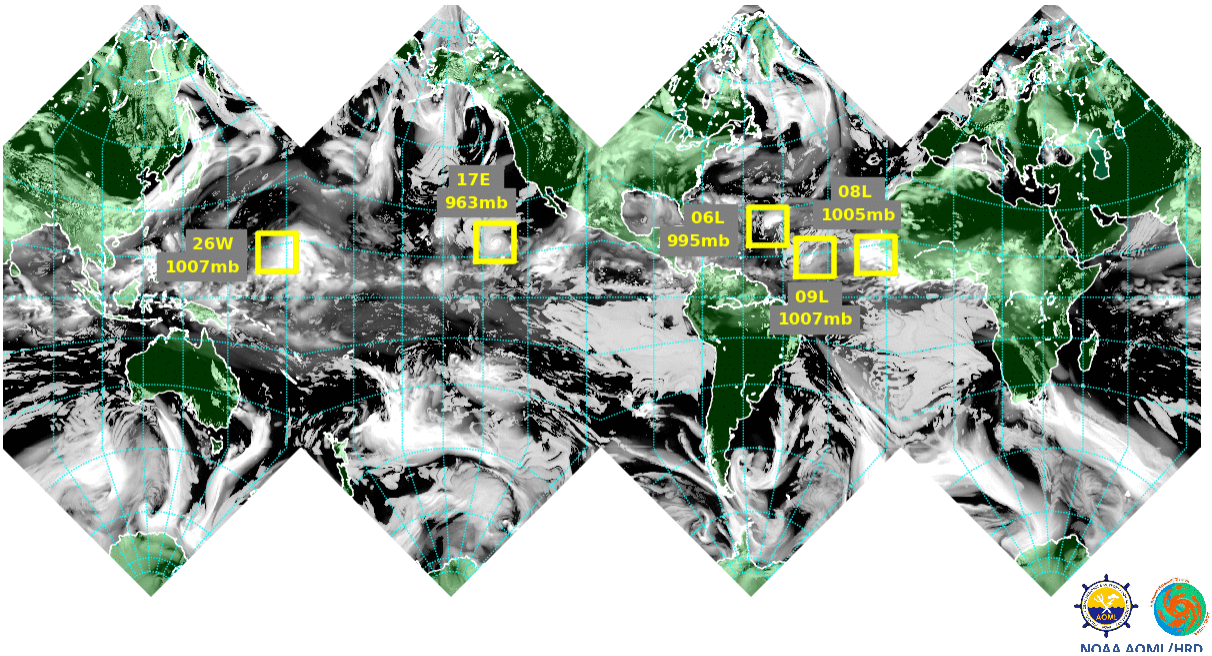
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
United States Department of Commerce

HURRICANE FORECAST IMPROVEMENT PROJECT

2019 HFIP R&D Activities Summary: Recent Results and Operational Implementation

May 2020

HFIP Technical Report: HFIP2020-1



NOAA AOML/HRD

Image on the cover page is a prototype showing how high-resolution nests can be moved seamlessly within the six faces of the FV3 cube sphere grid.

2019 HFIP R&D Activities Summary: Recent Results and Operational Implementation

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

May 2020

[1] Atlantic Oceanographic and Meteorological Laboratory (AOML)
[2] Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State University (CSU)
[3] Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder
[4] Developmental Testbed Center (DTC)
[5] Environmental Modeling Center (EMC)
[6] Global Systems Laboratory (GSL)
[7] Geophysical Fluid Dynamics Laboratory (GFDL)
[8] Hurricane Research Division (HRD)
[9] National Center for Atmospheric Research (NCAR)
[10] National Hurricane Center (NHC)
[11] Naval Research Laboratory (NRL)
[12] NOAA Environmental Software Infrastructure and Interoperability (NESII)
[13] Office of Science and Technology Integration (OSTI)
[14] Office of Planning and Programming for Service Delivery (OPPSD)
[15] Physical Sciences Laboratory (PSL)
[16] Pennsylvania State University (PSU)
[17] Science and Technology Corporation
[18] University at Albany, State University of New York
[19] University of Miami, CIMAS/AOML
[20] University of Oklahoma

This page is intentionally left blank.

Table of Contents

Executive Summary 7

1. Introduction 9

2. The Hurricane Forecast Improvement Program (HFIP) 9

3. The HFIP Baseline for measuring progress 10

4. The HFIP Model Systems 12

5. Operational HWRF and HMON systems (Stream 1) 13

- **a. HWRF System** 13
- **b. HMON System** 15
- **c. Challenging Case of Hurricane Dorian** 15

6. Operational Hurricane Guidance Improvements 16

- **a. Track Guidance** 16
- **b. Intensity Guidance** 17
- **c. Rapid Intensification/Weakening Prediction** 18

7. Development of the Hurricane Analysis and Forecast System (HAFS) 20

8. Important Stream-2 Results 22

- **a. HWRF Research Advances: Multiple, Storm-following, Two-way Interactive Telescoping Nests (Stream-2)** 22
- **b. HAFS Experimental systems (Stream-2)** 24

9. New Products, Tools, and Services at NHC 28

- **a. Operational and Real-Time Applications** 28
- **b. Display and Diagnostic Activities** 29
- **c. Experimental Projects** 29

10. Community Involvement 29

11. NOAA Federally Funded Opportunity (FFO) 31

12. Socio-economic Aspects of HFIP 31

13. Intensity Predictions: HFIP State-of-the-art and HAFS developments 33

14. Future direction of the HFIP 35

References 37

Appendix A: List of Acronyms 39

List of Figures

Figure 1: HFIP (a) Track and (b) Intensity Error Baseline and Goals. 11

Figure 2: Official NHC (a) Track errors (1960-2017) and (b) Intensity errors (1970-2017) in the AL basin..... 12

Figure 3: (a) HWRF intensity skill relative to Decay-SHIFOR for the 2011-2019 Atlantic seasons; (b) HWRF Track skill relative to CLIPER5 for 2011-2019 Atlantic seasons. 13

Figure 4: (a) HWRF intensity forecast skills over WPAC, (b) HWRF track forecast skills over WPAC. 14

Figure 5: Challenging forecast of Tropical Cyclone Dorian (a) Observed track and cones indicating the different phases where the operational models struggled example of (b) track forecast from various operational models during the initial phase and (c) track and intensity prediction during the second phase where Dorain rapidly intensified and made landfall in Bahamas. 15

Figure 6: Official track forecast skill in 2019 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. 17

Figure 7: Official intensity forecast skill in 2019 for the (a) Atlantic Basin (left) and (b) East Pacific Basin (right). 17

Figure 8: 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. 19

Figure 9: HFIP RI performance measure for 2018-19, presented in terms of skill relative to the HFIP baseline. CONI is meant to mimic the 2019 IVCN and comprises at least two of the models DSHP, LGEM, HWFI, HMNI, and CTCI. Results for 2019 are preliminary. 20

Figure 10: HWRF-B configured with three high-resolution movable multi-level nests following Hurricane Dorian, Tropical Storm Gabrielle, and Hurricane Juliette for a forecast initialized at 1800 UTC 03 September 2019. 23

Figure 11: Verification of a) mean track skill scores versus interpolated 2019 operational HWRF forecasts (HWFI; purple) and b) mean intensity skill scores versus HWFI for interpolated 2019 HWRF-B forecasts (HB9I; brown), interpolated GFS forecasts (GFSI; blue), interpolated HMON forecasts (HMNI; green), and interpolated CTCX forecasts (CTCI; gold). 24

Figure 12: The Atlantic static domain for the stand-alone regional version of HAFS (HAFS-SAR). 25

Figure 13: a) The 6 cubed-sphere tiles of the HAFS-globalnest global domain, with the Atlantic nest outlined in red. b) A zoomed-in view of the tile centered on the Atlantic, with the Atlantic nest outlined in red. 25

Figure 14: (a) Track forecast verification (mean error in n mi) for all 2019 cases from HAFS-SAR (light blue), HAFS-globalnest (red), HWRF (purple), HMON (green), and GFS (dark blue); (b) Intensity forecast verification (mean absolute error in kt) for all 2019 cases from HAFS-SAR (light blue), HAFS-globalnest (red), HWRF (purple), HMON (green), and GFS (dark blue). 26

Figure 15: Track forecasts of Hurricane Dorian through 00 UTC 04 September, 2019 for a) HAFS-SAR (126h forecasts) b) HAFS-globalnest (168h forecasts). 27

Figure 16: Examples of PPAV products and results from 2019 28

Figure 17: (a) Intensity diagnostic plot for the late-cycle forecast track guidance for Hurricane Michael, initialized at 18 UTC on 08 October 2018. (b) Lead time diagnostic plot for the late-cycle forecast track guidance for Hurricane Michael, initialized at 18 UTC on 08 October 2018. 30

Figure 18: The 5-day cone of uncertainty with black track line toggled off. 32

Figure 19: HWRF intensity skill relative to Decay-SHIFOR for 2011-2019 Atlantic seasons including 5-year and 10-year HFIP goals. 34

List of Tables

Table 1. HFIP RI performance measures baseline and target errors. 19

Table 2 HFIP Supported Projects from 2018-2020. 31

.

Executive Summary

This technical report describes the activities and results of the Hurricane Forecast Improvement Program (HFIP) occurred in 2019. The major development focus in 2019 was on building the next generation Hurricane model - Hurricane Analysis and Forecast System (HAFS), primarily for track and intensity predictions. This report will summarize the progress in 2019 including model developments and first year of progress made towards transforming into the next generation of HFIP.

In general, the 2019 hurricane season was not as busy as 2016-2018, yet it was a very challenging year for the numerical models. There were eighteen named storms formed, of which six developed into hurricanes, with three major hurricanes, Dorian, Humberto and Lorenzo. There were 93 occurrences of Rapid Intensification (RI) events as observed in best-track data.

The major highlights of 2019 were:

1. The Hurricane Weather and Research Forecasting (HWRF) model was upgraded to run at a horizontal resolution of 1.5 km near the storm region in 2018. This made HWRF the highest resolution hurricane model ever implemented for operations in the National Weather Service (NWS). However, due to the NCEP Central Operations (NCO) moratorium, HWRF was not operationally upgraded in 2019.
2. In the Atlantic Basin, HWRF had the best intensity skill at all lead times through day 5. In fact, HWRF had by far the best intensity skill of any dynamical model on day 3 and later, and only SHIPS had slightly better intensity skill than HWRF on day 4. In the west Pacific basin, Operational HWRF had the best intensity performance for all lead times. In the East Pacific basin, operational HWRF had the best intensity skill until day 2, and unusually high intensity errors beyond day 3. In the East Pacific, HWRF had good track skills for all lead times. However, in the Atlantic basin, HWRF had unusually large track errors very likely due to Hurricane Dorian.
3. In the Atlantic basin, track errors for Hurricanes in a Multi-scale Ocean-coupled Non- hydrostatic (HMON) model were comparable to HWRF. Both in the Atlantic basin and East Pacific basins, HMON had large intensity errors beyond Day 2 and extended lead times.
4. The basin-scale HWRF, a major HFIP investment that was continuously run in parallel under Stream 2, showed better skills in both track and intensity forecasting and was as successful as the operational HWRF for most of the Atlantic hurricanes in 2019. The HWRF-B moving nests are foundational for the development of National Oceanic and Atmospheric Administration (NOAA)'s FV3-based, next generation hurricane forecast system.
5. One of the major accomplishments for the season was the advancement and real-time testing of two basic configurations of FV3-based, NOAA's next-generation Hurricane Analysis and Forecast System (HAFS)- (i) high resolution regional stand-alone regional mode (HAFS v0.A) and; (ii) global model with a high resolution nest mode (HAFS v0.B). The results from the baseline versions demonstrated initial success of the model. During Hurricane Dorian, both HAFS v0.A and v0.B consistently followed the best track and more accurately predicted the right turn before the coasts of Florida.
6. Hurricane Dorian was particularly the most challenging case in the 2019 season. Most of the operational models did not accurately predict the track and intensity forecast. Nevertheless, HFIP stream 2 experimental models (HAFS v0.A & v0.B) came closest in predicting the track forecast. However, the major RI events were missed by most of the models. Guidance did not show strengthening of Dorian to its peak intensity of 160 kts within 24 hours. Rapid Intensification (RI) predictions continue to be a challenge for HAFS as well.

7. Supported by the NOAA Hurricane Supplemental projects, accelerated developments of HAFS are ongoing. 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 Unified Forecast System (UFS) developments.
8. Under the Weather Research and Forecasting Innovation Act including Sect. 104, HFIP will continue to address the 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, particularly with focus on RI 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). While significant progress were made, especially track and intensity predictions using the HWRF system, further improvements are necessary. The HAFS system is expected to address those new HFIP goals.

1. Introduction

This report describes the Hurricane Forecast Improvement Program (HFIP), its goals, proposed methods for achieving those goals, and recent results from the program, with an emphasis on recent advances in the skill of operational hurricane forecast guidance. The first part of this report is very similar to previous versions of the annual report, since it basically describes the background of the program. This year's version focuses upon capturing state-of-the-art HFIP modeling accomplishments during 2019's hurricane season, development of Hurricane Analysis and Forecasting System (HAFS), progress on the Rapid Intensification (RI) metrics, and future plans. For more background information, readers are referred to earlier reports available at: http://www.hfip.org/documents/.

2. The Hurricane Forecast Improvement Program (HFIP)

Twenty-seven named tropical storms and thirteen hurricanes crossed US coastlines from 2000-2010. The Hurricane Forecast Improvement Program (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:

- 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)[1] 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).

HFIP provides the unifying 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 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 the HFIP, NOAA continues to improve the accuracy of hurricane forecasts, with applied research using advanced computer models.

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

[1] 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).

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 Section 104 of the Weather Research Forecasting Innovation Act of 2017, is to advance an operational Hurricane Analysis and Forecast System (HAFS) at NOAA/National Weather Service (NWS). 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, and coupling to earth system components for high-resolution TC predictions within the outlined Next Generation Global Prediction System (NGGPS)/Strategic Implementation Plan (SIP) objectives of the Unified Forecast System (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 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 2019 season, an advanced version of Hurricane Weather and Research Forecasting (HWRF) model, called the Basin-Scale HWRF, and two preliminary versions of HAFS were tested near real-time in Stream 2 (see section 8 for results).

3. The 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, while for intensity the members were GFDI, DSHP, and LGEM[2] (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).

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

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.

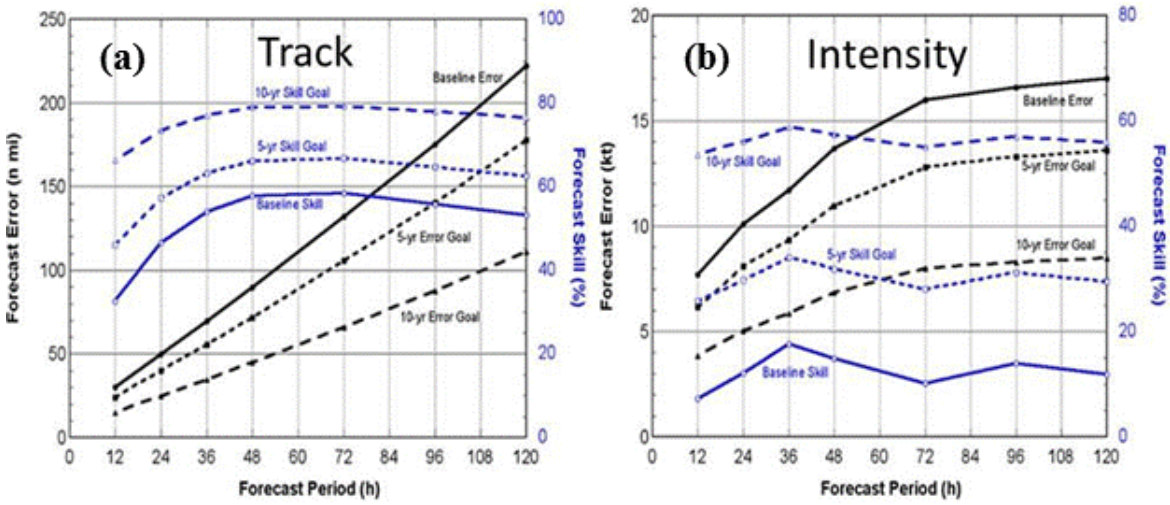


Figure 1: (a) Track and (b) Intensity Error Baseline and Goals, where the forecast errors are represented by black lines labeled on the left side of the graph, and the forecast skill is represented by blue lines labeled on the right side of the graph. Solid black lines represent baseline forecast errors,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

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 early 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. The 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. Fig. 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 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), and early forecast cycles of Florence (2017) 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.

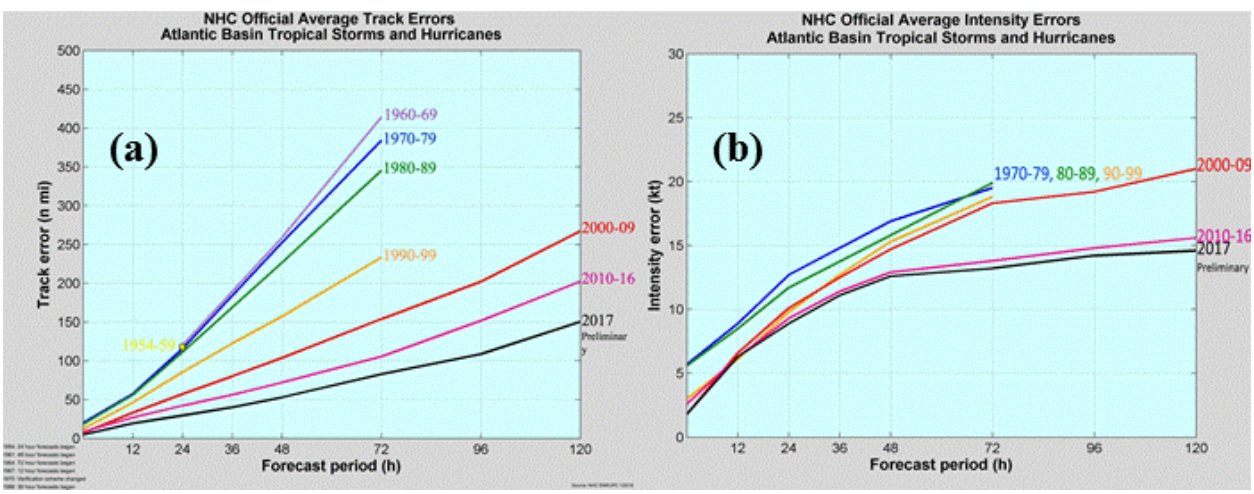


Figure 2: **Official NHC (a) Track errors (1960-2017) and (b) Intensity errors (1970-2017) in the AL basin.**

While significant track improvements have been achieved since the 60's, 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 10 years, the focus has been on improving intensity forecast, which for decades has significantly lagged behind track forecast. For that purpose, regional models with (two-way interactive) moving nests capable of resolving the inner core structure of hurricanes are usually used for intensity predictions. The domains of the hurricane regional models are usually larger than their CONUS counterparts. The HWRF and HMON that were developed during HFIP are prime examples. Track predictions from these regional models, especially HWRF, have been shown to improve, with larger domains (Zhang et. al., 2016; and Alaka et. al., 2017). 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. The envisioned Global nests embedded in the FV3 based HAFS, under development, is shown in the cover picture.

5. Operational HWRF and HMON systems (Stream 1)

a. HWRF System

One of the major accomplishments of HFIP has been the development of the storm-following, double-nested, high-resolution, HWRF model, and its transition to operations. A joint development between NOAA research and operations, with significant support from the Developmental Testbed Center (DTC), 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.

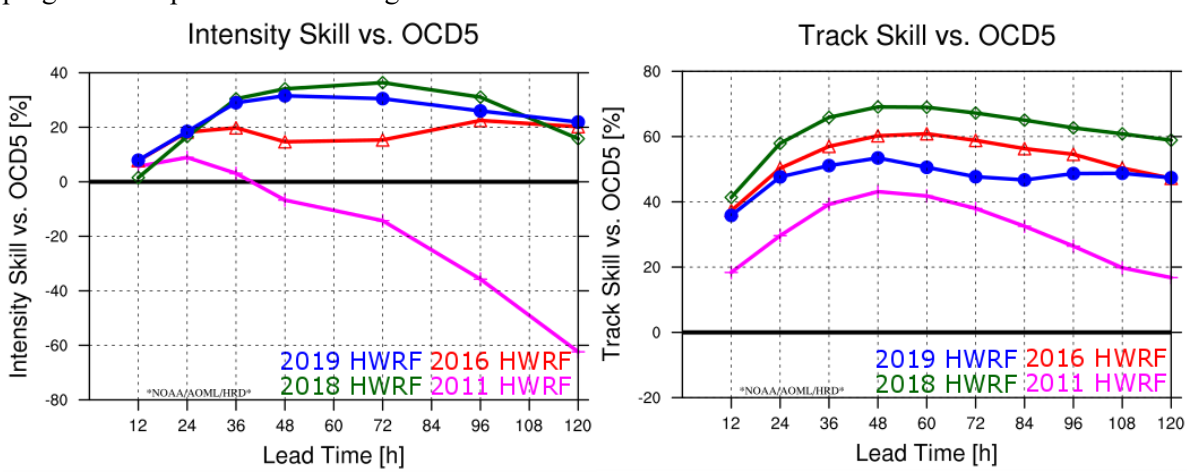


Figure 3: (a) HWRF intensity skill relative to Decay-SHIFOR for the 2011-2019 Atlantic seasons; (b) HWRF Track skill relative to CLIPER5 for 2011-2019 Atlantic seasons.

Figure 3a portrays the progress of HWRF in forecasting intensity, measured in terms of skill relative to Decay-SHIFOR. Through 2011, HWRF was operating with a single 9 km-resolution moving nest that could automatically track hurricanes (Gopalakrishnan et. al., 2006). In the next eight years (2012-2019), the HWRF system was upgraded considerably under HFIP year after 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.
- The year 2016 was the watermark year for 5-year improvements. New SAS and GFS-EDMF physics suites were implemented during this year.
- Supported by HFIP, a dramatically improved DA system was implemented in operational HWRF in 2017 (shown in Fig. 3a).
- 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 NWS.
- However, due to the NCEP Central Operations (NCO) moratorium, HWRF was not operationally upgraded in 2019.

Clearly, steep-step progress is being made under the HFIP with every yearly upgrade. HWRF had improved by about 40-60% from 2011-2018 (Fig. 3a). In fact, HWRF is 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. However, it should be noted, outlier events continue to impact HWRF performance from year to year. In 2019, Hurricane Dorain was a challenging forecast. In fact, HWRF performance slightly degraded in 2019 in terms of intensity skill when compared to 2018 (HFIP Annual Report, 2019) likely due to the challenging storms coupled with lack of model upgrades. However, HWRF was the best dynamical model for the intensity forecast over the Atlantic.

Figure 3b illustrates the track forecast skill relative to CLIPER from the HWRF system from 2011 to 2019. Although HWRF was initially developed for improving intensity guidance, the model is also used, complementary to the GFS, for providing track forecasts, as well. Track performance from HWRF has been constantly improving since 2011 until 2018 (Fig. 3b). However, as mentioned earlier, Dorian was a very challenging storm for HWRF in 2019. Substantial degradation of nearly 20% (HFIP Annual Report, 2019) was noted partly due to Hurricane Dorain. The parent GFS that drives the initial conditions was upgraded to FV3 core and that could influence the larger scales significantly in any regional models.

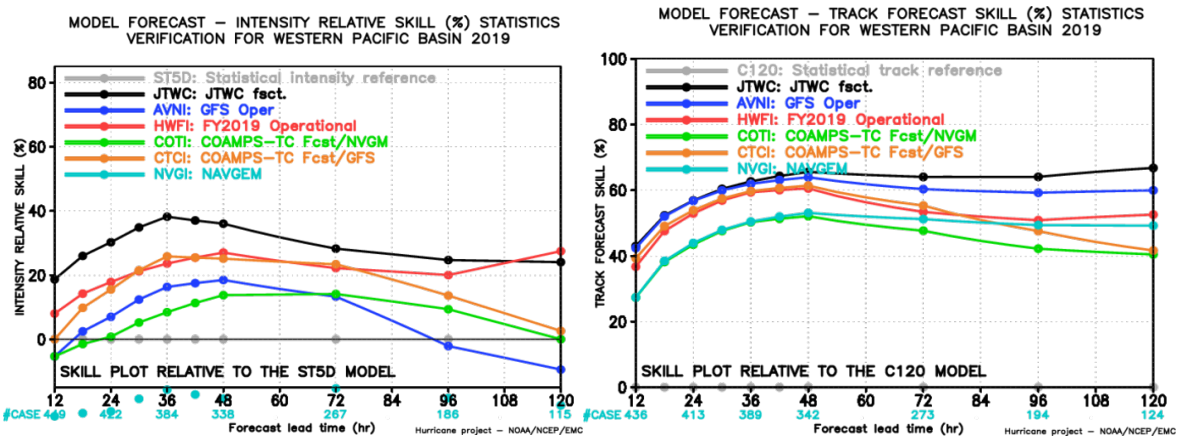


Figure 4: (a) HWRF intensity forecast skills over WPAC, (b) HWRF track forecast skills over WPAC.

In the Western Pacific basin, Operational HWRF has the best intensity performance at all lead times. HWRF intensity performance was exceptional along with COAMPS-TC (Fig. 4a). HWRF track forecast was the second best only behind GFS at all lead times and outperformed the COAMPS-TC at longer lead times (Fig. 4b).

b. HMON System

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

c. Challenging Case of Hurricane Dorian

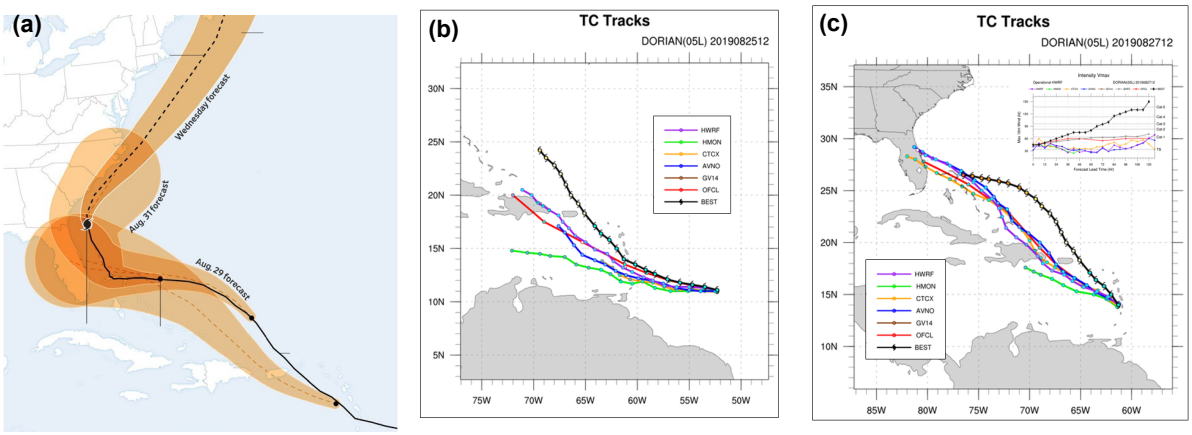


Figure 5: Challenging forecast of Tropical Cyclone Dorian (a) Observed track and cones indicating the different phases where the operational models struggled example of (b) track forecast from various operational models during the initial phase and (c) track and intensity prediction during the second phase where Dorain rapidly intensified and made landfall in Bahamas.

The Atlantic Hurricane Dorian was a forecasting challenge of the 2019 season. Dorian was formed on August 24, 2019 from a tropical wave in the Central Atlantic, and gradually strengthened as it moved toward the Lesser Antilles, becoming a hurricane on August 28, 2019. Rapid intensification occurred, and on August 31, Dorian became a Category 4 hurricane. On September 1, Dorian reached Category 5 intensity, with maximum sustained winds of 185 mph, and a minimum central pressure of 910 mb while making landfall in Elbow Cay, Bahamas. Dorian made another landfall on Grand Bahama several hours later. The ridge of high pressure steering Dorian westward collapsed on September 2, causing Dorian to stall just north of Grand Bahama for about a day. It is the strongest known tropical system to affect the Bahamas. A combination of cold water upwelling and an eyewall replacement cycle weakened Dorian to a Category 2 hurricane on the next day. On the morning of September 3, Dorian began to move slowly

towards the north-northwest. Dorian subsequently completed its eyewall replacement cycle and moved over warmer waters, regaining Category 3 intensity by midnight on September 5. In the early hours of September 6, Dorian weakened to Category 1 intensity as it picked up speed and turned northeast. Dorian would pick up speed and move northeast along the North Carolina coast September 6, moving just south of the Crystal Coast, clipping Cape Lookout and eventually making landfall at Cape Hatteras.

Figure 5 summarizes the complexity in Dorain's prediction and some of the model errors associated with track and intensity forecasts. The three cones in Figure 5a approximately indicate the behavior of the forecast models during different phases described above and the black line shows the actual track. The early forecasts showed the storm would stay weak and pass near or over Puerto Rico and/or Hispaniola. In reality, the center very likely re-developed further northeast (Avila et. al., 2020), and the observed track was outside the track forecast spread of operational models (Fig. 5b). The following forecasts had track as well as intensity forecast errors (Fig. 5c). None of the operational models, HWRF, HMON, SHIPS could forecast the RI predictions in several cycles, very likely due to large position errors. There were phase errors in many of these forecasts from numerical models, as well. For instance, when the storm stalled for 48 hours near the Bahamas, HWRF showed several cycles making landfall near West Palm Beach Florida. The following phase of rapid weakening and moving northeast along the North Carolina coast was not well captured by HWRF. This single storm accounted for a significant portion of the degradation in HWRF performance in the 2019 season. Nevertheless, as seen later, HFIP experimental systems, especially HAFS provided improved track forecast showing promise in NOAA's next generation effort.

6. Operational Hurricane Guidance Improvements

The 2019 Atlantic hurricane season was not as busy as 2016-2018. There were eighteen named storms formed, of which six developed into hurricanes, with three major hurricanes, Dorian, Humberto and Lorenzo. Dorian impacted the northern Bahamas as a category 5 hurricane producing catastrophic wind and surge damage while Hurricanes Humberto and Lorenzo affected Bermuda and the Azores, respectively. Hurricane Barry, a category 1 hurricane, affected the U.S. by making landfall.

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 and LGEM statistical models. 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, which is a linear average of the D-SHIPS and LGEM statistical models and the early versions of the HWRF, HMON regional models. IVCN runs when two or more of the above models (HWRF, HMON, D-SHIPS and LGEM) are available. We use IVCN as the basis for performance measures for RI predictions this year instead of individual model guidance from HWRF and HMON (section 6c).

a. Track Guidance

In 2019, official Atlantic track forecasts (Fig. 6a) were very skillful and close to the best-performing consensus aids - FSSE, HCCA and TVCA (Cangiolosi, 2019). EMXI was the best dynamical mode, but not as good as the NHC forecasts or consensus models. EGRI was the second best model, followed by AEMI and GFSI. NVFI, HMNI and HWFI trailed off.

In the eastern Pacific (Fig.6b), the consensus aids HCCA and TVCE led the way with the highest skill. NHC official forecasts were very good but a little less than the consensus models in the short term. EMXI was the best individual model, but less skill than the official forecast and consensus models. AEMI was close to, but not quite as good as, EMXI. EGRI was a strong performer through 96 h, with a second place model. GFSI, HWFI, HMNI were in the middle of the pack, just behind the best individual model. NVGI was completely trailed off.

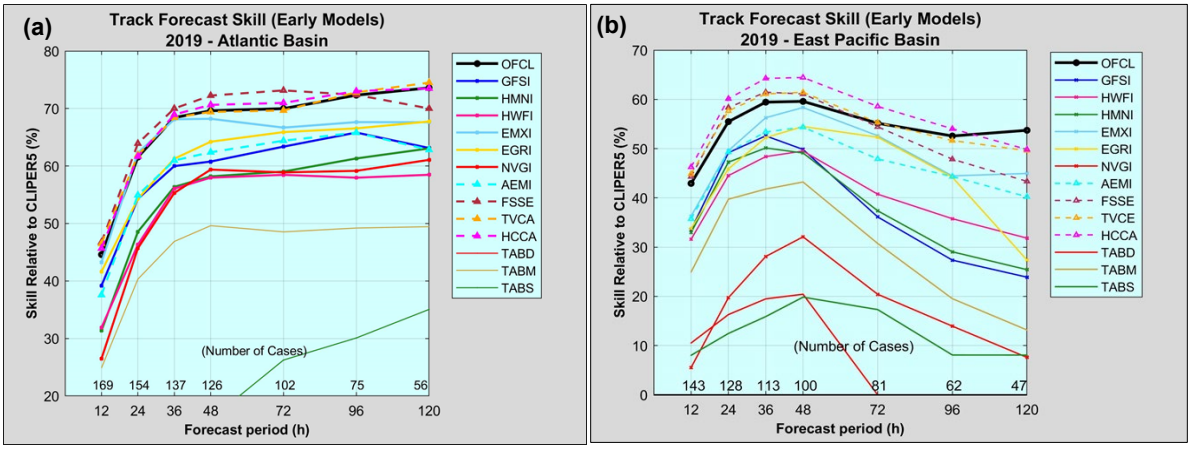


Figure 6: Official track forecast skill in 2019 for the (a) Atlantic (left) and (b) eastern Pacific (right) basins. Numbers immediately above the X-axis show the total number of cases covered by each data point.

b. Intensity Guidance

Intensity forecast verifications for the 2019 season are shown in Fig. 7. In the Atlantic basin (Fig. 7a), official forecasts were very skillful near the consensus aid. Among the consensus models, HCCA was the best model from 12h to 48h, while FSSE performance was best from 72h to 120h. HWFI was a strong performer, the best individual model at most lead times. HMNI was competitive with HWFI early, but trailed off after 48h. DSHP and LGEM were fair performers, but not as good as HWFI and consensus models. GFSI and EMXI had some skill, but not competitive with the remainder of the guidance.

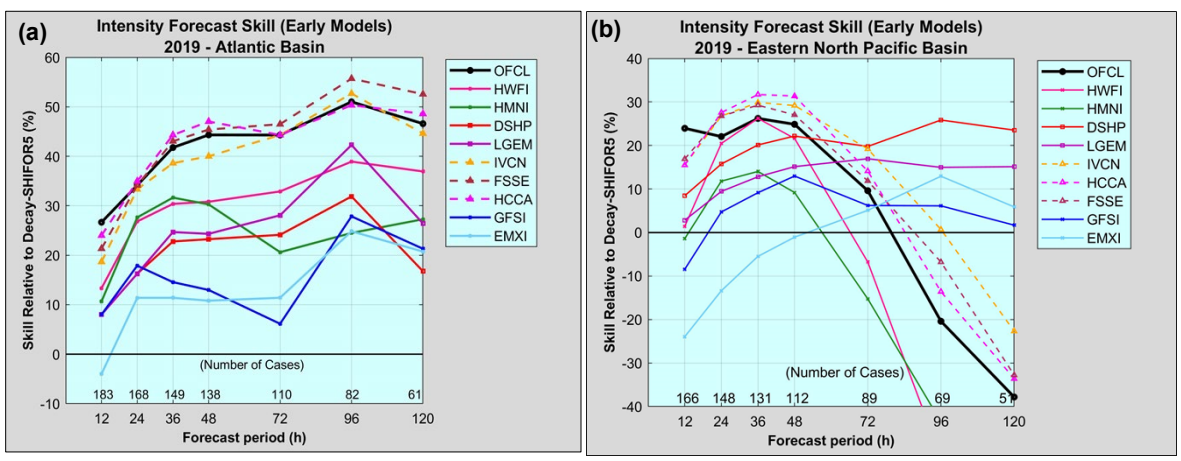


Figure 7: Official intensity forecast skill in 2019 for the (a) Atlantic Basin (left) and (b) East Pacific Basin (right).

In the eastern Pacific (Fig. 7b), official intensity forecast performance was better than the guidance early but not quite as good as the consensus aids. There were no skills at 96h and 120h. IVCN, FSSE, and HCCA started as best models early, but trailed off at longer lead times. HWFI was a strong performer through 48 h, but skill dropped off sharply after that. DSHP and LGEM were the only models that were skillful throughout and were best at 96h and 120h. GFSI and EMXI were not very skillful early, but had better guidance at the long leads. Many of the TCs over the eastern Pacific were short lived and posed a challenge to all models. (Cangiolosi, 2020).

c. Rapid Intensification/Weakening Prediction

One of the HFIP goals is to "reduce intensity forecast guidance errors by 50% for RI events". A number of specific options consistent with this language had previously been proposed, including one that was discussed in the 2018 HFIP Annual Report. After further discussion over the past year, however, a different proposal for measuring progress toward forecasting RI has been adopted. The new HFIP RI performance metric, baseline, and some preliminary results are discussed below.

The new 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.

By considering both RI cases occurring in the best track and the RI cases being forecast, the new 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.

A few additional remarks about the new metric are in order. First, while most HFIP performance measures have been applied to just the Atlantic basin, the rarity of RI events argued for a combined Atlantic/eastern Pacific evaluation to increase sample size. Second, because RI is assessed relative to the intensity 24-h prior, a baseline was not developed for 12-h forecasts. Finally, we anticipate that non-consensus forecasts (e.g., HWFI, OFCL) will be evaluated relative to the new RI baseline and target; it's hopefully clear that in such cases criteria (2) above would be applied to each of the models forming the homogeneous sample.

The values of the new RI baseline are presented in Table 1 and Fig. 8. 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

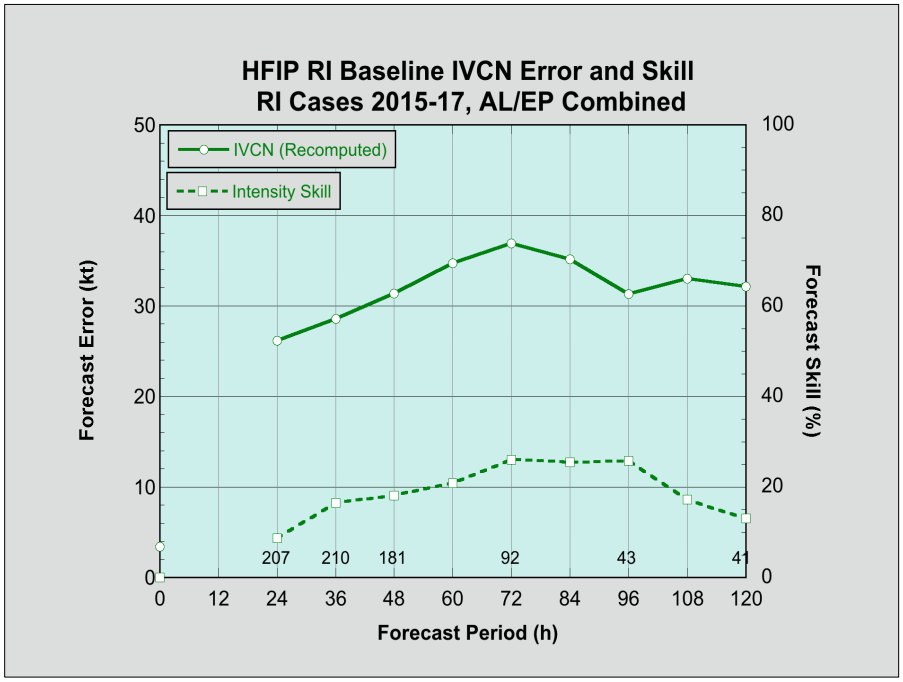


Figure 8: HFIP RI performance measures baseline errors and skill. Baseline errors are the mean absolute errors over the period 2015-17 for the Atlantic and eastern North Pacific for the variable consensus comprising at least two of the models DSHP, LGEM, GHMI, HWFI, and CTCI. Skill values are computed relative to OCD5.

Figure 9 shows how the RI intensity metric has performed over the past two seasons (note that 2019 results are preliminary since the season's best tracks have not been finalized). The consensus forecast shown here (CONI) is defined as at least two of DSHP, LGEM, HWFI, HMNI, and CTCI – the 2019 composition of IVCN. Results are presented in terms of skill relative to the HFIP RI baseline errors from Table 1. It is seen that RI forecasts from the consensus were more successful in 2018 than in 2019, and it's fair to infer that there will be high year-to-year variability in the performance of this metric, in part due to the scarcity of RI events. Taking the two years together does show overall improvement over the baseline of at least 10% through 72 h, where the sample is relatively large. Errors were worse than the baseline at 96 and 120 h, however. Examination of the individual components of the consensus shows that LGEM and HMNI were relatively poor performers, while CTCI and HWFI were the models primarily responsible for the improvements relative to the baseline.

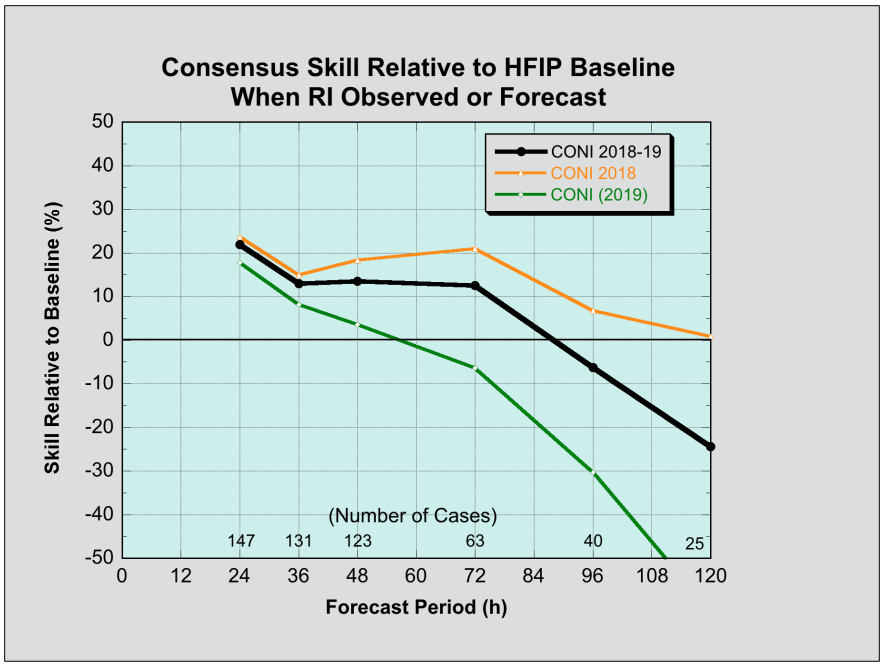


Figure 9: HFIP RI performance measure for 2018-19, presented in terms of skill relative to the HFIP baseline. CONI is meant to mimic the 2019 IVCN and comprises at least two of the models DSHP, LGEM, HWFI, HMNI, and CTCI. Results for 2019 are preliminary. Number of cases along the bottom of the diagram refers to the combined Atlantic and eastern Pacific 2018-19 sample.

7. Development of the Hurricane Analysis and Forecast System (HAFS)

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

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

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

While active developments of the HAFS system enlisted above are ongoing, for the 2019 season, two preliminary configurations of the HAFS system were run under Stream-2. Some of the preliminary results, especially related to track predictions of Dorain where the operational models struggled, showed promise in the next generation hurricane forecast system i.e. HAFS.

8. Important Stream-2 Results

a. HWRF Research Advances: Multiple, Storm-following, Two-way Interactive Telescoping Nests (Stream-2)

Although the operational HWRF system is improving intensity forecast skill, it is currently configured with only one set of high-resolution nests (i.e., it is storm-centric). This is not ideal for forecasting storm-storm interactions, storm-environment interactions, or TC genesis. Further, the limited size of the outermost operational HWRF domain may limit the improvement of forecast skill beyond five days, a major goal of next-generation numerical weather prediction efforts. HWRF's configuration poses many challenges for producing a large-scale analysis because its outermost domain moves from one forecast to the next, which is incompatible with current data assimilation software. All of these points may represent impediments to further advances in hurricane forecast guidance from dynamical models. For this reason, a Basin-scale HWRF (HWRF-B) 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 simultaneous TC forecasts at high resolution, and also serves as a prototype for the development of multiple moving, multi-level nests within the global model. HWRF-B allows for advanced data assimilation evaluations given its static outermost domain, and has already shown promise in observing system experiments (OSEs) and observing system simulation experiments (OSSEs). HWRF-B has been a collaborative effort between AOML and NCEP.

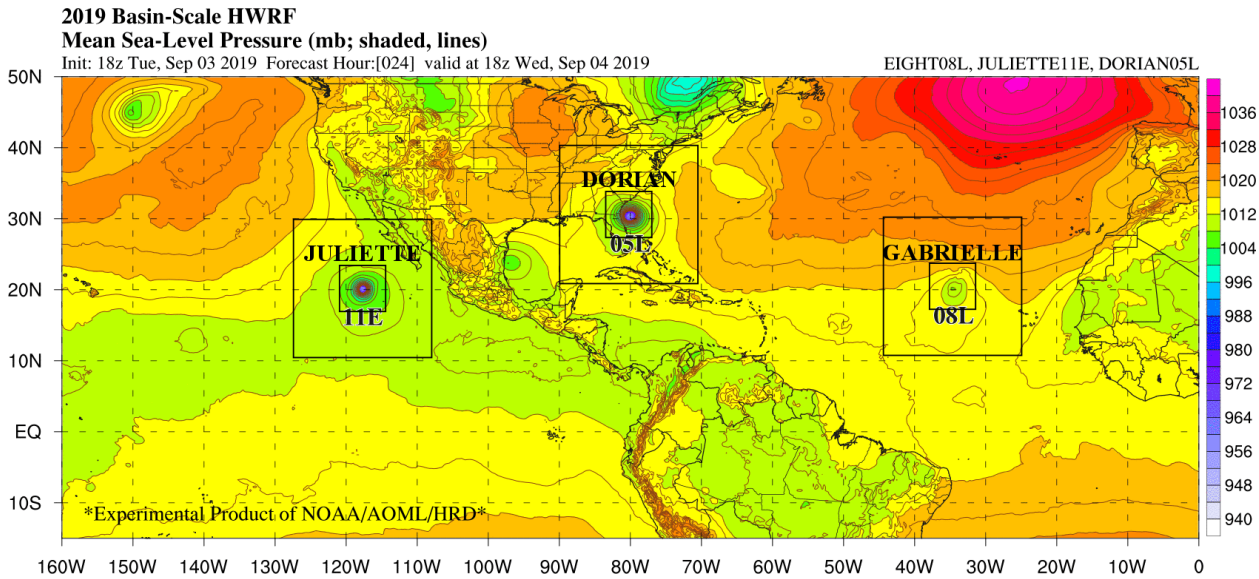


Figure 10: HWRF-B configured with three high-resolution movable multi-level nests following Hurricane Dorian, Tropical Storm Gabrielle, and Hurricane Juliette for a forecast initialized at 1800 UTC 03 September 2019.

The experimental HWRF-B model was an HFIP real-time demonstration for the seventh year, in parallel with operational hurricane models, during the 2019 North Atlantic and eastern North Pacific hurricane seasons. HWRF-B was run for the first time on NOAA's Weather and Climate Operational Supercomputing System (WCOSS), moving it a step closer to operational transition. A multi-storm coupler was developed 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 at a time at high resolution (~1.5 km) during the 2019 real-time demonstration; offseason experiments are now tracking up to five TCs and more are possible with expected increases to high performance computer resources. In 2019, HWRF-B assimilated the exact same satellite, ground-based, and aircraft observations as the operational system, and, after the season, the identical self-cycled data assimilation software used for the operational system was transitioned to HWRF-B and expanded to multiple TCs. With high cyclone activity in both basins, HWRF-B produced forecasts for multiple TCs in > 90% of its simulations. For example, high-resolution nests in HWRF-B tracked Hurricane Dorian, Tropical Storm Gabrielle, and Hurricane Juliette for a forecast initialized at 1800 UTC 03 September 2019 (Fig. 10). HWRF-B was at least 20% better than the operational HWRF for intensity predictions of Hurricane Dorian at most forecast lead times. HWRF-B outperformed the operational HWRF in 2019, especially when multiple TCs were active in the North Atlantic and eastern North Pacific basins. Seasonal error statistics show that the 2019 HWRF-B (HB9I) had more skillful track and intensity forecasts than the operational HWRF (HWFI) at most (especially longer) lead times (Fig. 11). 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. In contrast, the operational HWRF is capable of tracking only one storm at high resolution per forecast and may misrepresent storm-storm interactions critical for intensity forecasts (e.g., Hurricane Dorian). The HWRF-B moving nests are foundational to the next

generation HAFS. The advanced and well-evaluated nesting technique is being transitioned to the FV3 Unified Forecast System (FV3-UFS).

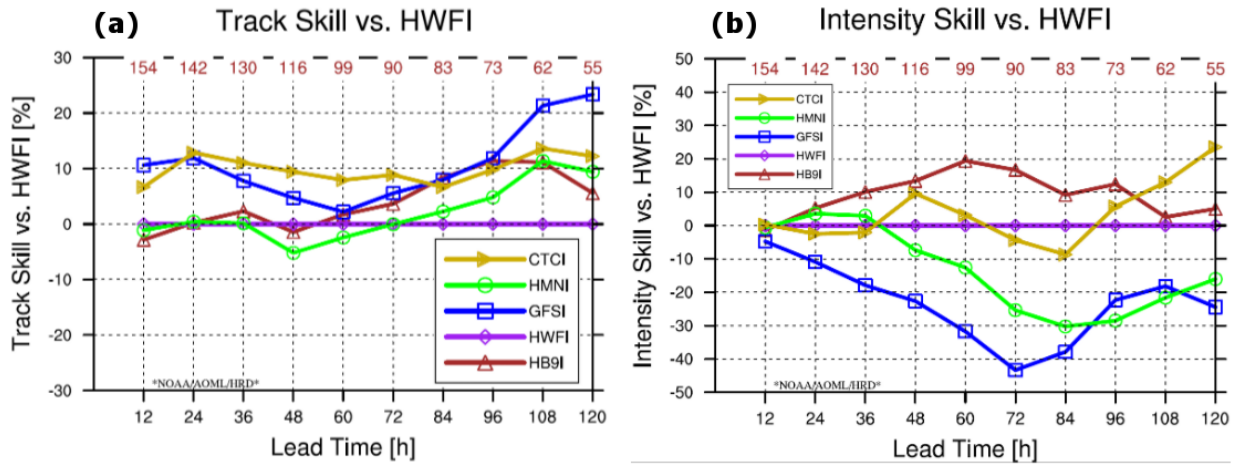


Figure 11: Verification of a) mean track skill scores versus interpolated 2019 operational HWRF forecasts (HWFI; purple) and b) mean intensity skill scores versus HWFI for interpolated 2019 HWRF-B forecasts (HB9I; brown), interpolated GFS forecasts (GFSI; blue), interpolated HMON forecasts (HMNI; green), and interpolated CTCX forecasts (CTCI; gold).

b. HAFS Experimental systems (Stream-2)

As a part of Stream 2, two preliminary versions of the HAFS systems, the global with one static nest at 3 km resolution covering the Atlantic basin (HAFS0.B) and another version of a stand-alone version of a regional configuration at uniform 3-km resolution (HAFS0.A) was run during the 2019 hurricane season. This system neither had moving nests nor had the advanced inner core DA packages critical for improved intensity and structure predictions.

HAFS0.A configuration uses the stand-alone-regional (SAR) configuration of FV3, with a large static domain covering the north Atlantic at ~3-km grid spacing (Figure 12) that gets initial conditions from the GFS analysis, and boundary conditions in one-way feedback from the operational GFS every three hours. HAFS0.B configuration is based on the global-nested configuration of FV3GFS (Harris and Lin, 2013), with a large static nest over the Atlantic at ~3-km grid spacing (similar to Hazelton et al. 2018b) inside of a global forecast (at ~12-13 km grid spacing). The nest covers the entire Atlantic TC genesis/development region from off the coast of Africa, through the Atlantic, Caribbean, and Gulf of Mexico (Figure 13). This configuration of HAFS allows for two-way feedback between the global and static nested domains. This version was called HAFS-globalnest (or HAFSB). For 2019, both versions of HAFS were "cold starts", meaning that they were initialized directly from the 13-km global GFS analysis. The development of a data assimilation system for HAFS, including assimilation of radar and other relevant TC data, is an ongoing project. Both versions of HAFS use 64 vertical levels on the sigma-pressure hybrid coordinate with the lowest model level at about 25 m above the surface and the top level at 0.2 hPa. Physics parameterizations are similar to those in the operational FV3GFS, with modifications to the surface drag (Bender et al. 2007) and PBL physics (Wang et al. 2018) for more realistic results in TC environments. The horizontal advection scheme used in HAFS-SAR and the nest domain of HAFS-globalnest is more diffusive than operational FV3GFS to keep the high resolution forecast stable. HAFS-SAR was run for 126 hours, and HAFS-globalnest for 168 hours. This version of HAFS did not include an ocean coupling component, a capability currently in development.

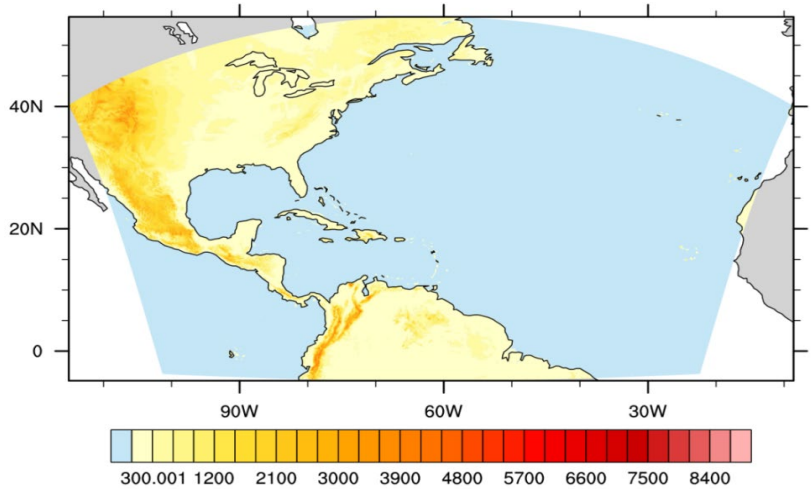


Figure 12: The Atlantic static domain for the stand-alone regional version of HAFS (HAFS-SAR).

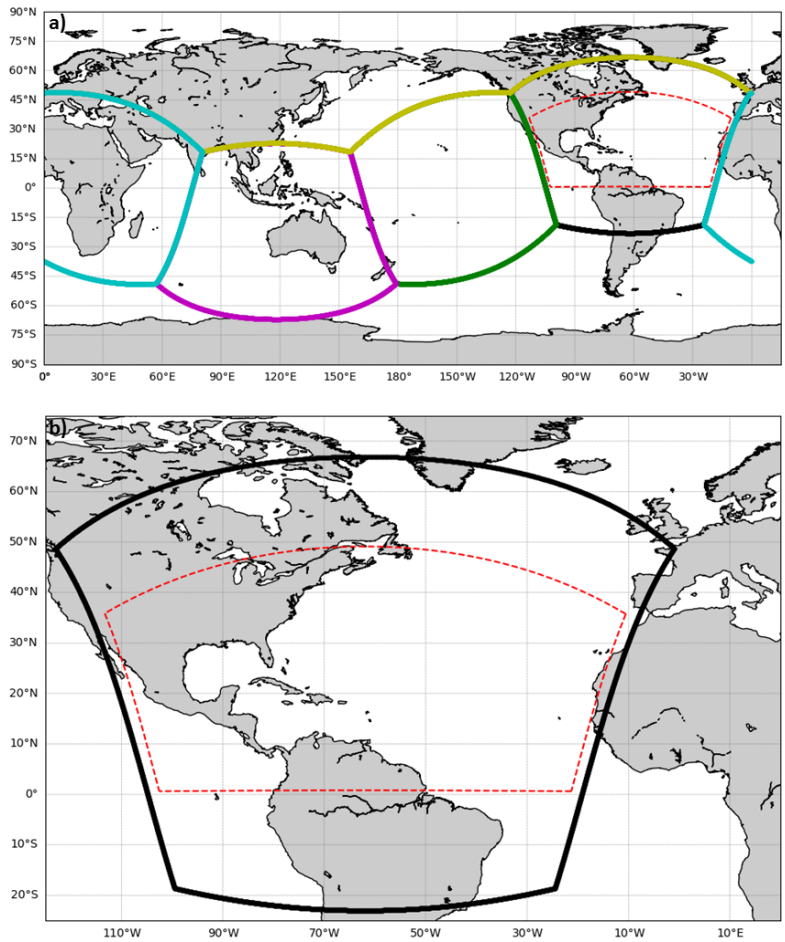


Figure 13: a) The 6 cubed-sphere tiles of the HAFS-globalnest global domain, with the Atlantic nest outlined in red. b) A zoomed-in view of the tile centered on the Atlantic, with the Atlantic nest outlined in red.

Figure 14a shows the track forecast verification from HAFS and other models for the 2019 Atlantic Hurricane season. Both versions of HAFS performed well, with lower errors at all hours out to Day 5 than the GFS, HWRF, and HMON. There is about a 20% improvement over HWRF in terms of track skill throughout 5 days forecasts from both HAFS versions. Most of the track forecast improvement over GFS/HWRF/HMON is from the cross-track component. This is a promising result that demonstrates the potential for HAFS to further improve TC track prediction.

Figure 14b shows the intensity forecast verification from HAFS and other models for the 2019 Atlantic Hurricane Season. The results are a little more mixed than those for track forecast. While the intensity forecast errors were lower than the operational GFS at all forecast hours, demonstrating the value of the high-resolution nest for TC prediction, the errors were higher than the current operational hurricane models (HWRF and HMON) at early lead times. Both HAFS versions have comparable or smaller intensity error than HWRF/HMON for weak storms (initial intensity less than 50 kts) while generally underpredict the intensity for strong storms (initial intensity more than 50 kts). This points to a need for ongoing development of HAFS, including improvements to horizontal and vertical resolution, model physics, and data assimilation to improve intensity forecasts, especially during challenging cases like rapid intensification. All of these topics (and others) are the focus of ongoing research and development efforts.

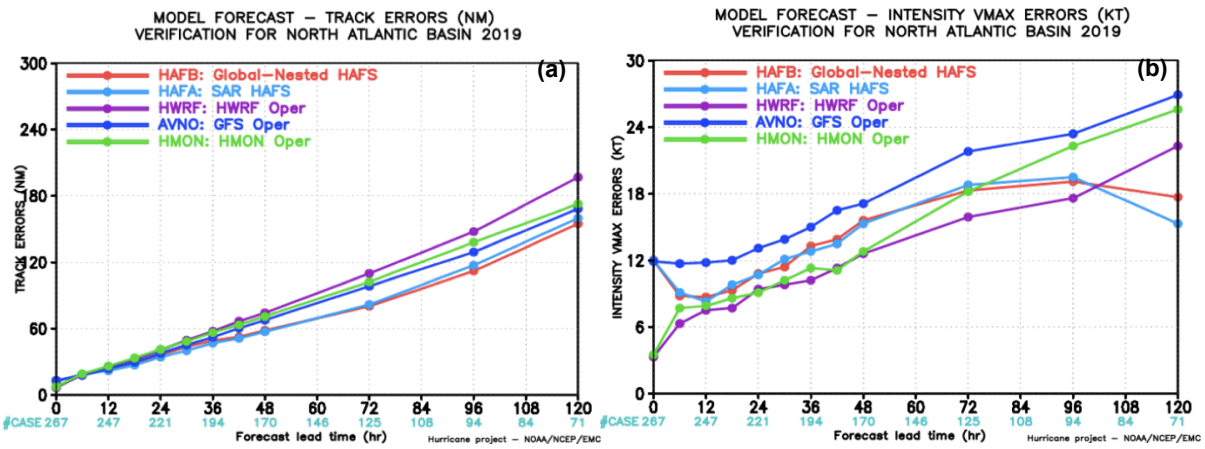


Figure 14: (a) Track forecast verification (mean error in n mi) for all 2019 cases from HAFS-SAR (light blue), HAFS-globalnest (red), HWRF (purple), HMON (green), and GFS (dark blue); (b) Intensity forecast verification (mean absolute error in kt) for all 2019 cases from HAFS-SAR (light blue), HAFS-globalnest (red), HWRF (purple), HMON (green), and GFS (dark blue).

One of the biggest successes for HAFS was the forecasts for Hurricane Dorian. Both versions of HAFS were consistently showing Dorian turning to the East of Florida, while the GFS and other operational models predicted a Florida landfall (Fig. 15).

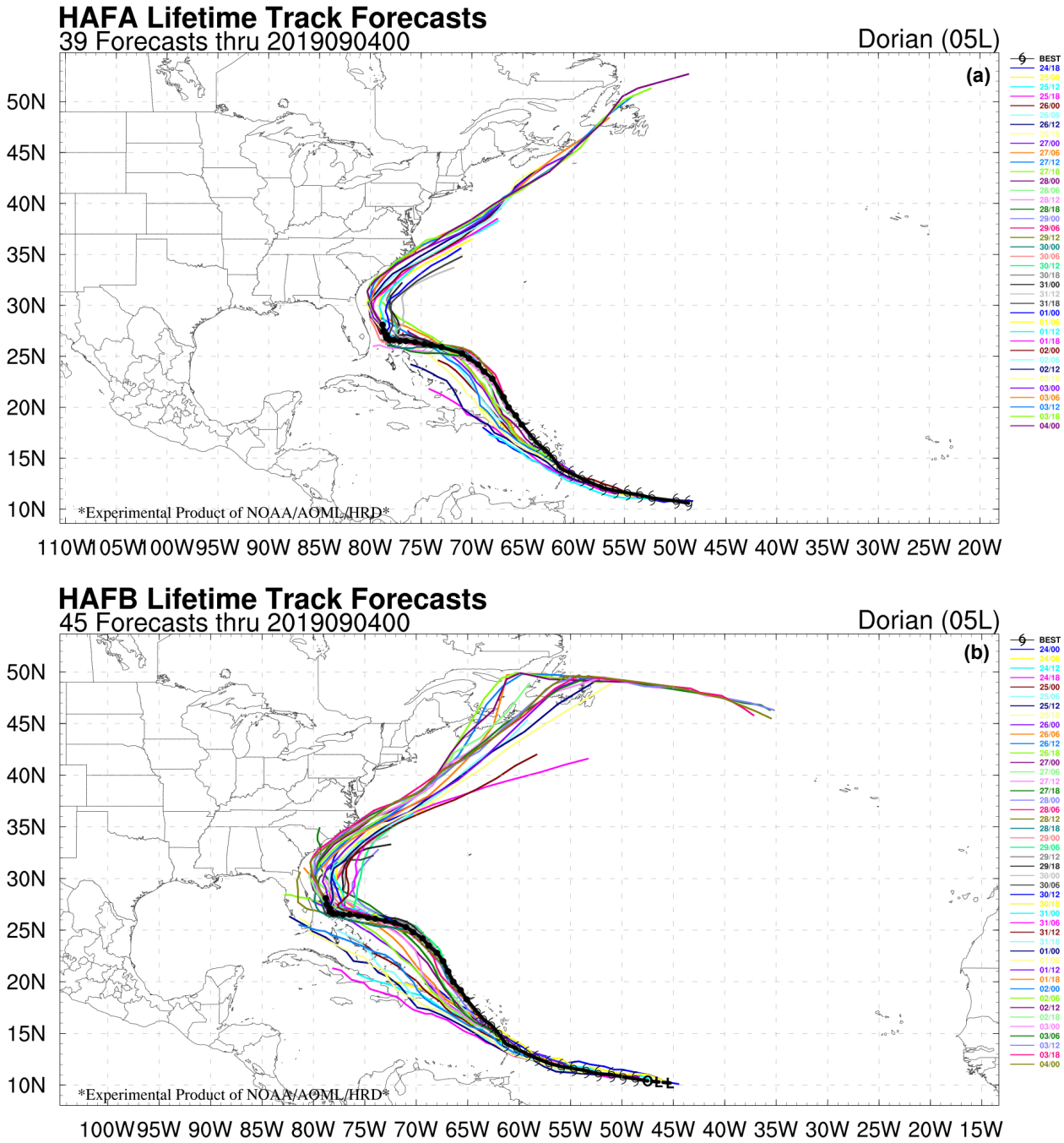


Figure 15: Track forecasts of Hurricane Dorian through 00 UTC 04 September, 2019 for - a) HAFS-SAR (126h forecasts) b) HAFS-globalnest (168h forecasts).

9. New Products, Tools, and Services at NHC

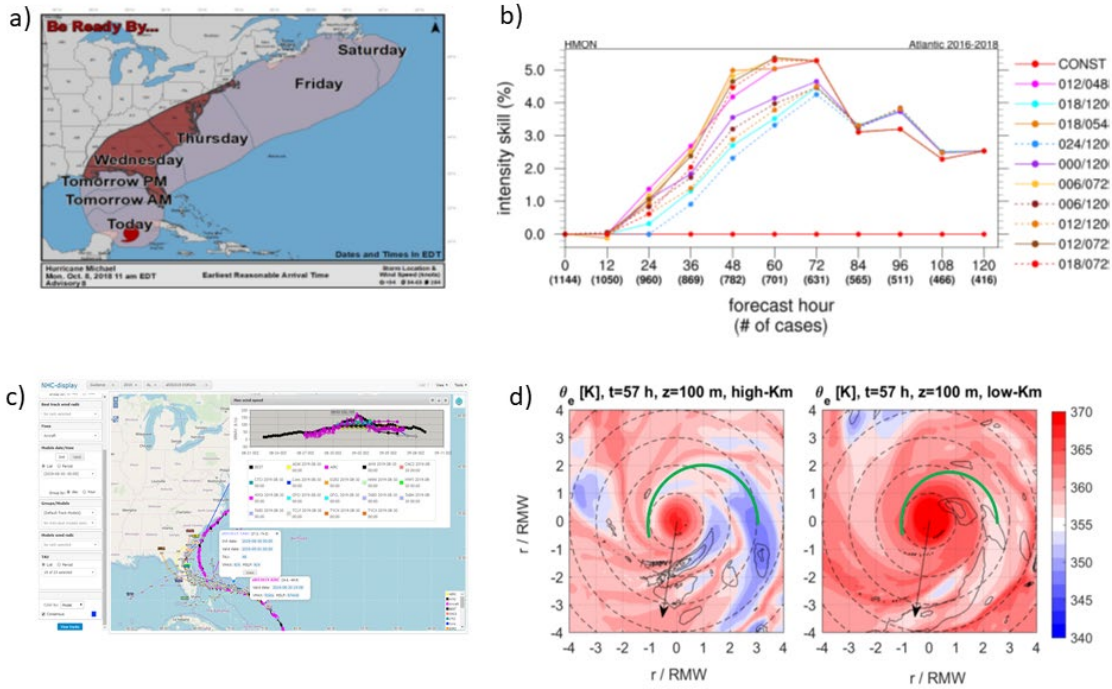


Figure 16: Examples of PPAV products and results from 2019: a) Refined version of the "Be Ready By" graphic; b) Intensity skill of HMON forecasts with different interpolation schemes. The higher lines represent more skill (better forecasts); c) HFIP NHC diagnostic display example from Hurricane Dorian showing track and intensity information; d) Boundary layer entropy in HWRF forecasts with different vertical eddy diffusivity. Entropy is much smaller in the high eddy diffusivity forecast (left) than the low diffusivity forecast (right). The low diffusivity forecast more accurately captured the rapid intensification of Hurricane Earl.

a. Operational and Real-Time Applications

Great strides were made in 2019 toward the improvement of operational tools and real-time diagnostic applications for hurricane forecasting. Improvements were made to the output graphics and operational stability of the HFIP Corrected Consensus Approach (HCCA). Automatic training with each forecast cycle was also built in, and new models were added to the track and intensity consensus. The operational track and intensity consensus aids, TVCN and IVCN, were also optimized for 2019, as was the interpolator used to create operational guidance from late-arriving model output at the NHC (Fig. 16b). Cluster-guidance based on ensemble output was developed for 34- and 50-kt wind radii forecasts. A rapid-intensification model, DTOPS (Deterministic to Probabilistic Statistical RI Index), ran operationally for the first time in 2019 and showed promising results. Another new statistical model, the Neural Network Intensity Combination (NNIC) was tested in 2019 and shows promise at further improving consensus-based guidance.

HFIP also supported improvements to operational infrastructure and real-time prototype products. This included updates to the ATCF used by NHC forecasters to compose operational advisory packages. A wind speed probability model-based gridded forecast (WTCM) was tested in real-time in 2019 for evaluation by WFOs and NHC/TAFB. The WTCM converts the NHC text forecast to a gridded format and includes the variation in surface roughness due to land surface type and the feedback in 2019 led to several improvements to the parametric wind model. In AWIPS, a hazard recommender tool was tested based on the wind speed probability model. This tool is expected to help forecasters with future warning decision making and ensure consistent products and messaging across multiple local forecast offices during hurricane landfall events. Enhanced verification of watches and warnings and the wind speed

probabilities were developed to support that effort. Finally, a simplified version of the time of arrival product, the "Be Ready By" graphic was refined (Fig. 16a) and available for briefings and improved decision support services at the NHC.

b. Display and Diagnostic Activities

The HFIP community continued to make progress toward improving model diagnostics and visualization in 2019. At the NHC, work was done to create an extended HURDAT-2 database. New capabilities and optimizations were also built into the NHC verification code, including a tie-in to environmental parameter diagnostic files from the SHIPS model. NCAR developed new diagnostic products to help visualize ensemble-based rapid intensification forecasts using parameters such as intensity and pressure. This concept will be applied to other diagnostics such as wind shear or sea surface temperature. National Center for Atmospheric Research (NCAR) also continued to support web-based tools including the NHC Display and Diagnostic system and provided a number of enhancements to that visualization tool (Fig. 16c). The HFIP Products website and ESRL TC Tracks pages are other web diagnostic tools that were supported and improved in 2019. Hurricane Research Division (HRD) used diagnostics to guide HAFS development by placing model output in a framework that could be directly compared to radar observations. HRD also developed ways to visualize the impact of boundary layer parameterizations on hurricane structure and intensification in the HWRF model (Fig. 16d).

c. Experimental Projects

In addition to the progress outlined above, three projects supported by the Hurricane Supplemental began work in 2019. Recent aircraft reconnaissance observations in hurricanes have shown an apparent high bias from the Stepped Frequency Microwave Radiometer (SFMR) instrument. One project has begun to re-examine the calibration based on the inclusion of more recent data and re-examine data co-location and adjustment algorithms. Another project is looking at new ways to examine hurricane forecasts through the use of 3-D visualization software to improve real-time diagnostics and facilitate post-storm analysis. The third project is focused on extending the NHC's existing forecast guidance to 7 days and modernizing the code to help with long-term maintenance and development. This work is critical as the NHC examines the utility of extending its forecasts to 7 days.

10. Community Involvement

Research to Operations (R2O) was one of the initial goals of the WRF program and is supported by HFIP in developing a repository for a community-based hurricane modeling system, which ensures the same code base can be used for research and in operations. During 2009-2016, both the EMC and the DTC worked to update the operational version of HWRF from version 2.0 to the community version of HWRF, version 3.9a. The 3.9a version made the operational model completely compatible with codes in community repositories, allowing researchers to access the operational codes. Hence, the improvements in HWRF, developed by the research community, were easily transferable into operations. DTC has played a significant role to help the HWRF community by conducting HWRF training sessions twice per year from 2010-2018, two of which were international. In addition, twelve Community Workshops on topics ranging from physics, observations, ensemble product development, satellite DA, to social science were conducted. In July 2018, the code version of the HWRF system v4.0a was available for the HWRF community. Since then DTC has continued to provide user support. Apart from US, there are about thousand HWRF model users in about 198 countries[3]. 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

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

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.

HFIP Ensemble Diagnostic Products: NCAR focused on a new initiative to develop diagnostic products that will help forecasters understand why particular models are forecasting rapid intensification (RI) while others do not. The initial efforts have resulted in the development of several new prototype visualizations, which show forecasted trajectories of the various forecast aids, with each forecast aid's trajectory colored according to the value of some diagnostic parameter. Possible diagnostic parameters could include environmental conditions (e.g., 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 prototype visualizations use parameters available from the ATCF a-decks, such as intensity, minimum sea level pressure, and forecast lead-time. Figure 17a shows an example visualization for Hurricane Michael, where the diagnostic parameter is the forecast aid's predicted intensity (VMAX). Figure 17b shows a similar diagnostic prototype visualization, but where the diagnostic is forecast lead-time. The trajectories of each forecast aid are colored by its forecast maximum intensity, using color bins corresponding to the categories of the Saffir-Simpson Hurricane Scale.

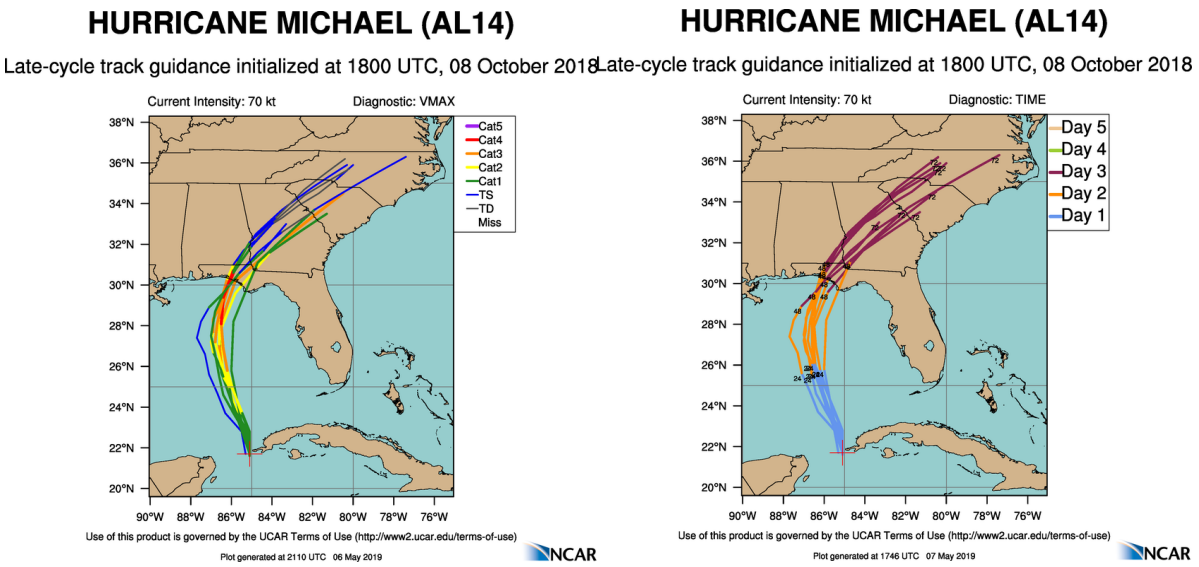


Figure 17: (a) Intensity diagnostic plot for the late-cycle forecast track guidance for Hurricane Michael, initialized at 18 UTC on 08 October 2018. (b) Lead time diagnostic plot for the late-cycle forecast track guidance for Hurricane Michael, initialized at 18 UTC on 08 October 2018.

NCAR NHC Display System: The new hurricane display and diagnostic capabilities allow forecasters and research scientists to more deeply examine the performance of operational and experimental models. The system is built upon modern and flexible technology, including OpenLayers Mapping tools and an efficient MySQL database. New technologies developed this past year include an advanced tool for editing the hurricane fix-position database (F-deck) and the best-track database (B-deck). The F-deck editing tool allows users to add or edit the estimated location of hurricanes using fixed-position information from aircraft analysis, radar, satellite, microwave, and scatterometer observations. A wind radii tool has been added to view wind radii graphics by 34kt, 50kt, 64kt wind thresholds. The display system also calculates and displays derived fields using GFS model output including wind shear, moisture, and precipitable water. An example of the display system capabilities is shown in Figure 16c.

11. NOAA Federally Funded Opportunity (FFO)

The following Table provides a list of projects supported by HFIP during 2018-2020.

Table 2 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

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 usually extend well beyond the bounds of the cone, the use of the cone in this way is disconcerting and potentially dangerous.

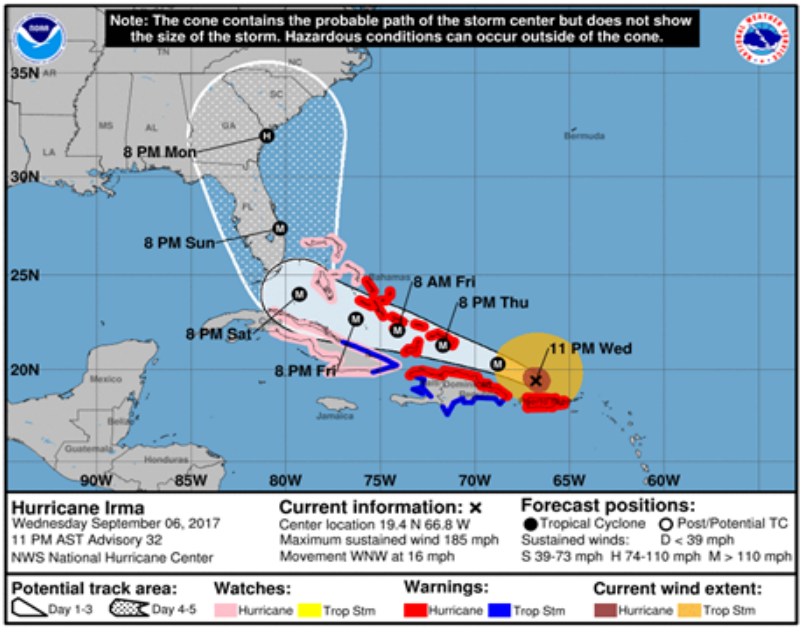


Figure 18: **The 5-day cone of uncertainty with black track line toggled off.**

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.

Though the cone may have complexities, conveying a tropical cyclone's uncertainty is very important. Scientific advancements continue to increase forecast accuracy, but uncertainties remain due to observational and modeling limitations of the steering currents surrounding a tropical cyclone. These limitations can lead to longer-term deviations of the storm's center from the official forecast track, or even normal shorter-term wobbles around the official forecast track. In addition, the hazards associated with tropical cyclones, including wind, storm surge, heavy rainfall, and tornadoes, can extend well away from the storm's center. These realities mean that people in both the direct and indirect path of the center of a tropical cyclone may need to prepare for the associated hazards. 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. Importantly, this "heads up" function is equally vital for people on land as compared to people over water, such as mariners.

Despite the cone's complexity, the cone remains one of the most 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 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 goal of this research is to synthesize what prior research and NWS assessments reveal about the cone and its interpretation and use. NOAA would also like to understand how embedded the cone of uncertainty is in stakeholder decision-making, and what those decisions and implications look like. The study will include a literature review of the general public, broadcast meteorologists and emergency

managers interpretation and understanding of the cone, including key decisions and decision-times of emergency managers based on the information provided by the cone. The study will also focus on the use of the cone by the wider, less-studied user base beyond emergency managers, including but not limited to utility companies, the tourism sector, transportation (including airlines, rail), marine (including fishing, cargo, ports, etc.), finance and insurance companies, military, etc. To the extent possible, the use study should include both domestic and as appropriate international users.

13. Intensity Predictions: 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, which are not possible in global models 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 intensity forecasting (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 (Fig. 2b). HWRF has improved by about 40-60% since 2011 over the Atlantic basin (Fig. 3). 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 nest capable of better resolving eyewall convection and scale interactions, inner core DA technique, 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.

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 (Fig.10). 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 (Fig.10). Transitions of this multiple moving nested HWRF to next generation global and regional HAFS system within the unified forecast system is underway and is expected to provide another step in improvements to the hurricane prediction capacity in NOAA.

Other noteworthy developments under HFIP were the Ensemble HWRF system, HyCOM-Wavewatch coupled HWRF system and fully cycled Basin-Scale HWRF (Gopalakrishnan, 2018 HFIP report). The systems are actively used, respectively, for research, especially related to understanding RI predictions (e.g., Leighton et. al., 2018) and for improving satellite data assimilation (Poterjoy et. al., 2019). The HCCA model has been another major achievement for the HFIP program (Simon et. al., 2018). Leveraging the success of HWRF and the capacity that was built under HFIP, a second high resolution hurricane model for intensity predictions, HMON, was developed by scientists at EMC and AOML under the Sandy Supplemental effort. The HMON replaced the legacy GFDL hurricane model and is now an integral part of NHC's forecast consensus.

These developments and T&E would not be possible without the support of the 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[4].

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 a 20% reduction in both storm track and intensity errors in the numerical guidance. Yet, as shown in Figure 19, although HWRF has improved 40-60% in intensity predictions, those improvements have only met 50% of the targeted HFIP intensity goals (i.e. we are only half way through!). Part of the reason may be associated with the lack of progress with dynamical guidance until 2012. In fact, until 2012 intensity predictions lagged even the baseline (Figure 19) primarily set on statistical-dynamical models (SHIPS and LGEM). Also it is not clear at this time what are the limiting factors of intensity predictability. In fact, the same kind of analysis as shown in Figure 19 was carried out by calculating the median errors. The errors shown in Figure 19 were further reduced significantly indicating that the outlier events are the ones that drive the larger errors. Some sustained HFIP research is recommended in this area especially with the next generation hurricane prediction system.

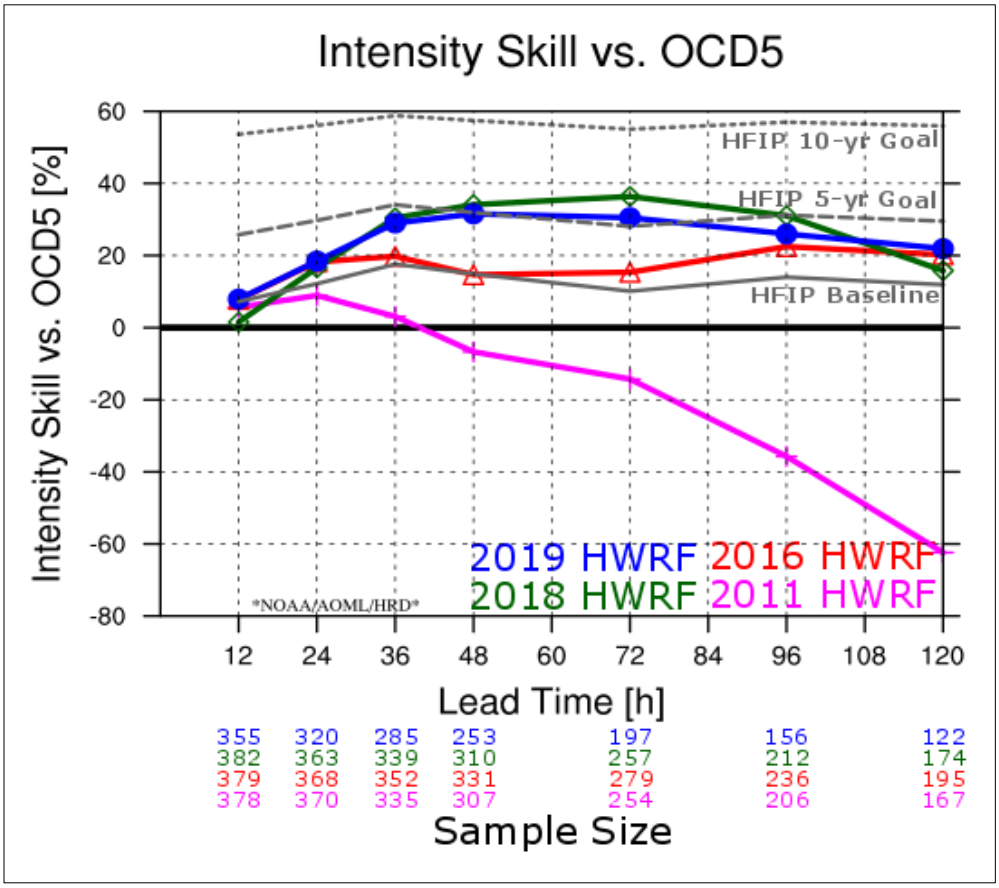


Figure 19: HWRF intensity skill relative to Decay-SHIFOR for 2011-2019 Atlantic seasons including 5-year and 10-year HFIP goals.

[4] http://www.hfip.org/documents

Improving RI (increase >30 kt intensity change in 24 hours) forecasts is one of the highest priorities for HFIP and was recognized as the most challenging aspect of TC research. The lack of improvement in the RI forecast skill is rooted in our lack of understanding of when and how RI occurs in different environmental conditions and the historic inability of dynamical models to accurately predict not only convection in the hurricane core, but also large scale environmental factors such as shear and moisture that produce an RI event (Chen and Gopalakrishnan, 2015; Leighton et. al., 2018). While HWRF is able to capture some of the complex cases of RI in a highly sheared environment (e.g., Hurricane Michael, 2019), storm to storm (e.g., Hurricane Patricia, 2015, Dorain, 2019) and cycle to cycle inconsistency (e.g., Hurricanes Harvey, 2018) makes the RI prediction still a very elusive problem. As shown in section 6c (Fig. 9), even multi-model consensus consisting of statistical models, where HWRF is the main dynamical model, have shown season to season variability. In fact, the 2019 season shows negative skill for most of the forecast time (Fig. 9). Some sustained HFIP research with HAFS is recommended in this area.

Supported by the NOAA Hurricane Supplemental projects, 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. Under the Weather Research and Forecasting Innovation Act including Section 104, HFIP will continue addressing the 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, 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). While significant progress were made with, especially track and intensity predictions using the HWRF system, further improvements are necessary. The HAFS system is expected to address those new HFIP goals.

14. Future direction of the HFIP

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 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:

1. Improving the prediction of rapid intensification and track of TCs;
2. Improving the forecast and communication of surges from TCs; and
3. 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.

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

1. Reduce forecast guidance errors, including during rapid intensification, by 50 percent from 2017;
2. Produce 7-day forecast guidance as good as the 2017 5-day forecast guidance;
3. Improve guidance on pre-formation disturbances, including genesis timing, and track and intensity forecasts, by 20 percent from 2017; and
4. 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.

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 the 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 accelerate 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 which has been recently allocated to support 2019-2022 HFIP activities. 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 the HFIP goals.

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.

Avila, L. A., S. R. Stewart, R. Berg, and A. B. Hagen, (2020): National Hurricane Center Tropical Cyclone Report, Hurricane Dorian (24 August – 7 September 2019), National Hurricane Center, April 2020. [Available online https://www.nhc.noaa.gov/data/tcr/AL052019_Dorian.pdf]

Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, (2007): The operational GFDL coupled hurricane–ocean prediction system and a summary of its performance. Mon. Wea. Rev., 135, 3965–3989.

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.

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., (2019): National hurricane center 2019 forecast verification (preliminary). The 11th HFIP Annual Rev. Conf. presentation, Nov. 4, 2019, Miami, FL. [Available online at http://www.hfip.org/events/annual_meeting_Nov 2019_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. *Wea. Forecasting,* **9**(2):209-220, doi:10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2.

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://www.hfip.org/documents/HFIP_AnnualReport_FY2018.pdf]

Harris, L. and S.J. Lin (2013): A Two-Way Nested Global-Regional Dynamical Core on the Cubed-Sphere Grid. Monthly Weather Review, 141, 283-306, https://doi.org/10.1175/MWR-D-11-00201.1.

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.

Leighton, H., S. Gopalakrishnan, J.A. Zhang, R.F. Rogers, Z. Zhang, and V. Tallapragada (2018): Azimuthal distribution of deep convection, environmental factors and tropical cyclone rapid intensification: A perspective from HWRF ensemble forecasts of Hurricane Edouard (2014). Journal of the Atmospheric Sciences, 75(1):275-295. https://doi.org/10.1175/JAS-D-17-0171.1.

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

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

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

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

Wang, W., J.A. Sippel, S. Abarca, L. Zhu, B. Liu, Z. Zhang, A. Mehra, and V. Tallapragada, (2018): Improving NCEP HWRF simulations of surface wind and inflow angle in the eyewall area. Wea. Forecasting, 33, 887–898.

Zhang, X., S.G. Gopalakrishnan, S. Trahan, T.S. Quirino, Q. Liu, Z. Zhang, G. Alaka, and V. Tallapragada, (2016): Representing multiple scales in the Hurricane Weather Research and Forecasting modeling system: Design of multiple sets of movable multilevel nesting and the basin-scale HWRF forecast verification. Weather and Forecasting, 31(6):2019-2034. https://doi.org/10.1175/WAF-D-16-0087.1.

Appendix A: List of Acronyms

AEMI	GEFS with 6 hour interpolation
AOML	Atlantic Oceanographic and Meteorology Laboratory
AVNI	GFS with 6 hour interpolation
AWIPS	Advanced Weather Interactive Processing System
CCPP	Common Community Physics Package
CLIPER	Climate and Persistence model
CMC	Canadian Meteorological 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
DA	Data Assimilation
DTC	Developmental Testbed Center
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	UKMO model, subjective tracker, 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
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
HNMMB	Hurricane Non-hydrostatic Multi-scale Model on B-grid
HPC	High Performance Computing
HRD	Hurricane Research Division
HWHI	Basin-scale HWRF with 6 hour interpolation
HWMI	HWRF Ensemble Mean Forecast Interpolated Ahead 6 hour
HWRF	Hurricane Weather and Research Forecasting
HWFI	HWRF with 6 hour interpolation
HYCOM	HYbrid Coordinate Ocean Model
IOOS	Integrated Ocean Observing System
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
OAR	Oceanic and Atmospheric Research
OFCL	Official National Hurricane Center Forecast
OSEs	Observing system experiments
OSSE	Observing system simulation experiments
POD	Probability of Detection
POM	Princeton Ocean Model
RI	Rapid Intensification
RW	Rapid weakening
SAR	Stand Alone Regional
SFMR	Stepped-Frequency Microwave Radiometer
SIP	Strategic Implementation Plan
SHIFOR	Statistical Hurricane Intensity Forecast
SHIPS	Statistical Hurricane Intensity Prediction System
SPICE	Statistical Prediction of Intensity from a Consensus Ensemble
SPIN-UP	Slang terminology for vortex acceleration and/or initialization
SPIN-DOWN	Slang terminology for vortex deceleration and/or termination
SREF	Short Range Ensemble Forecast
SST	Sea surface temperature
SSS	Sea surface salinity
TAB	Trajectory And Beta (TAB) model for trajectory track using GFS input
TC	Tropical Cyclone
TVCA	Track Variable Consensus of at least two of AVNI, EGRI, EMXI, NGPI, GHMI, HWFI forecasts
TVCE	Variable Consensus of AVNI, EGRI, EMXI, NGPI, GHMI, GFNI, HWFI Model Track Forecasts
TVCI	Variable Consensus of AVNI, EGRI, EMXI, NGPI, GHMI, GFNI, HWFI Model Track Forecasts (6 hour interpolation)
TVCN	Track Variable Consensus
UFS	Unified Forecast System
UKMI	United Kingdom Meteorological Office model with 6 hour interpolation
UW4I	University of Wisconsin's Non-hydrostatic Modeling System (4 km)
UWNI	UW-NMS with 6 hour interpolation (UWNI)

UW-NMS	University of Wisconsin Non-hydrostatic Modeling System
WMO	World Meteorological Organization
WRF	Weather Research & Forecasting
WFO	Weather Forecast Office