NOAA

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
United States Department of Commerce

HURRICANE FORECAST IMPROVEMENT PROJECT

2018 HFIP R&D Activities Summary: Recent Results and Operational Implementation

May 2019

HFIP Technical Report: HFIP2019-1

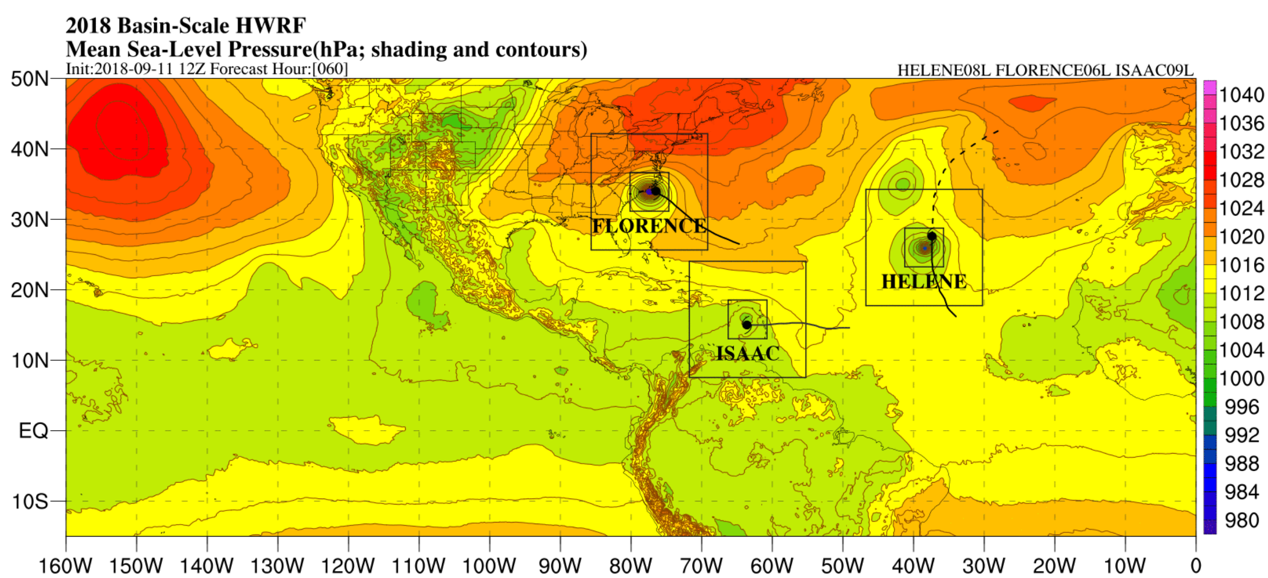

Image on the cover page is of Hurricanes Florence (06L, left), Issac (09L, center) and Helene (08L, right) from the Basin Scale HWRF initialized on 12Z September 11, 2018, at forecast hour 06z.

2018 HFIP R&D Activities Summary: Recent Results and Operational Implementation

S. Gopalakrishnan[7], F. Toepfer[11], Dorothy Koch[11], S. Upadhayay[15], M. DeMaria[9], F. Marks[7], E. N. Rappaport[9], Avichal Mehra[4], V. Tallapragada[4], G. Alaka Jr.[7], M. Bender[6], L. Bernardet[3, 5,19], J. Cione[1], M. Biswas[8], M. Brennan[9], J. Cangialosi[9], J. Doyle[10], J. L. Franklin[9], L.Gramer[17], G. Halliwell[1], A. Hazelton[7], J. S. Hilderbrand[12], E. Kalina[3,5,19], H. S. Kim[4], P. Kucera[8], N. Kurkowski[11], N. Lett[15], Bin Liu[4], P. McCaslin[5], E. Mifflin[15], M. Morin[6], J. Moskaitis[10], K. Musgrave[2], L. Nance[3], K. Newman[3], M. Onderlinde[9], A. Penny[9], A. Simon[9], J. Sippel[7], H. Tolman[11], R. Torn[16], X. Wang[18], J. Whitaker[13], H. Winterbottom[4], D. A. Zelinsky[9], F. Zhang[14], X. Zhang[7], Z. Zhang[4] and Lin Zhu[4].

May 2019

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

This page is intentionally left blank.

Table of Contents

List of Figures

List of Tables

This page is intentionally left blank.

Executive Summary

This technical report describes the activities and results of the Hurricane Forecast Improvement Project (HFIP) that occurred in 2018. In general, the 2018 hurricane season was representative of above normal activity over the Atlantic. There were fifteen named storms formed, of which eight developed into hurricanes, with two major hurricanes, Florence and Michael, reaching Category 3 or higher.[1] There were at least 20 occurrences of Rapid Intensification (RI)[2] events. The majority (15) of the RI cases were from Hurricanes Florence and Michael. The other five were from Hurricanes Beryl (1 event), Chris (2), and Oscar (1), and Tropical Storm Nadine (1). Some of the RI events, for example, in Tropical Cyclones Beryl, Chris, and Oscar, were very brief and difficult to predict. Meanwhile the East Pacific, with 23 named storms, had its fourth-most active season on record.

This report outlines HFIP, how it is organized, its goals, its models, and results. HFIP is organized around two streams: Stream-1: Operational model development and, Stream-2: HFIP experimental models, which test and evaluate new techniques and strategies for numerical model forecast guidance, prior to testing for possible operational implementation. Stream-2 also tests techniques that cannot be tested on current operational computers due to size and time requirements, but can be tested on HFIP's High Performance Computing (HPC) Center located in Boulder, CO (also referred to as Jet). HPC research studies look ahead, to possible future operational computational capabilities. As in the previous year, the major developmental focus in 2018 was on Operational Hurricane Weather Research and Forecasting (HWRF) and Operational Hurricanes in a Multi-scale Ocean-coupled Non-hydrostatic (HMON) regional models for track and intensity predictions.

The major highlights of 2018 were:

1. The HWRF model was upgraded to run at a horizontal resolution of 1.5 km near storm region. This would make HWRF the highest resolution hurricane model ever implemented for operations in the National Weather Service (NWS). Other HWRF upgrades consisted of physics advancements, continued improvements to the initialization package, system enhancements, and improved products. The HMON was upgraded in 2018 to run with ocean coupling.

2. In the East Pacific, HWRF was the best dynamical model with the lowest intensity errors. In the Atlantic basin, HWRF was the best dynamical model with the lowest intensity errors prior to Day 3. During that period, its intensity errors were comparable to those of the official forecasts from the National Hurricane Center (NHC). A significant portion of the intensity error from HWRF beyond Day 3 in the Atlantic basin was associated with one Tropical Cyclone, namely, Isaac.

3. HWRF performed well for both of the major landfalling hurricanes, namely, Florence and Michael. Some cycles of HWRF forecasts captured the RI of Hurricane Michael at least 4 days in advance. It should be noted that the storm developed in a hostile environment of shear exceeding 20-25 knots, where RI predictions can be a challenge.

4. For the first time, during hurricane Lane, P3 aircraft were flown in the Central Pacific, for assimilating inner core winds in the HWRF model. Post-analysis of model forecasts indicated an average of 20% track improvements, with a maximum of 35% at 96 hours, with the inclusion of the tail doppler radar data for initializing the HWRF system.

5. The HFIP Corrected Consensus Approach (HCCA) model has been a major achievement for the HFIP program. Further improvements to the model were made in 2018, including the migration of the code to the NWS operational supercomputing framework, the addition of Central Pacific

---

[1] https://www.nhc.noaa.gov/text/MIATWSAT.shtml

[2] RI for a tropical cyclone is defined as an increase in the maximum sustained winds of at least 30 kt in a 24 h period. This goal for HFIP also applies to rapid weakening (RW) - a decrease of 25 knots in 24 hours.

storms for the Central Pacific Hurricane Center (CPHC) area of responsibility, real-time updates to the training dataset, and the evaluation of the HMON as an additional input.

6. The basin-scale HWRF, a major HFIP investment that was continuously run in parallel under Stream 2, showed superior skills for Isaac intensity forecasting and was as successful as the operational HWRF for all other Atlantic hurricanes in 2018. Environmental Modeling Center (EMC) and Hurricane Research Division (HRD) are working to test the basin-scale HWRF system for possible operational implementation in 2020.
7. Post-analysis of the 2018 season showed that the basin-scale HWRF, not only covers a domain encompassing of both Atlantic and East Pacific, but is also capable of tracking simultaneously all the hurricanes in the domain at a horizontal resolution of 1.5 km (-vs. operational HWRF that can track only one hurricane in a forecast) captured the storm-storm interactions between Hurricanes Isaac, Florence and Helene much better (cover page image), demonstrating a viable pathway for hurricane moving nest in the Next Generation Global Prediction System (NGGPS).
8. Although the yearly HWRF upgrades demonstrated further reduction of errors, both on the track and intensity predictions, demonstrating the positive impacts of model upgrades since 2012 in predicting average and well behaved tropical cyclones, extreme events, namely, brief yet rapid intensification of hurricanes Beryl, Chris, and Oscar, and rapid weakening of TC Isaac, continue to pose forecasting challenges. Predicting RI of TCs remains the most significant challenge for forecasting. Additional, sustained HFIP research is recommended in this area.
9. Transitions of the multiple-moving-nested HWRF (basin-scale HWRF) to NOAA's Finite Volume Cubed-Sphere (FV3) based Hurricane Analysis and Forecasting System (HAFS) for tropical cyclone (TC) predictions within National Centers for Environmental Prediction (NCEP)'s Unified Forecast System (UFS; NGGPS implementation)[3] is underway, and expected to provide further improvements to NOAA's Next Generation hurricane prediction capacity.

[3] https://www.weather.gov/sti/stimodeling_nggps_implementation

1. Introduction

This report describes the Hurricane Forecast Improvement Project (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 2018's hurricane season, progress on the Rapid Intensification (RI) problem, 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 Project (HFIP)

Twenty-seven named tropical storms and thirteen hurricanes crossed US coastlines from 2000-2010. The Hurricane Forecast Improvement Project (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)[4] 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.

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.

A major component of Stream-2 is an Experimental Forecast System (EFS) that HFIP runs each hurricane season. The purpose of the EFS (also known as the Demonstration Project) is to evaluate the strengths and weaknesses of promising new approaches that are testable only with enhanced computing capabilities. The progress of Stream-2 work is evaluated after each season, to identify techniques that

---

[4]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).

appear particularly promising to operational forecasters and/or modelers. These potential advances can be blended into operational implementation plans through subsequent Stream-1 activities, or further developed outside of operations within Stream-2. Stream-2 models represent cutting-edge approaches that have little or no track record and, therefore, are not used by NHC forecasters to prepare their operational forecasts or warnings. Nevertheless, most of the operational HWRF advancements, including the high-resolution nests, appropriate physics, and data assimilation (DA) upgrades originated from Stream-2 work.

The new HFIP Strategic Plan detailing the specific research, development, and technology transfer activities necessary to sustain HFIP in response to Section 104 of the Weather Research Forecasting Innovation Act, was approved by NOAA and awaits Congressional approval. The major goals of the act will be addressed through the development of a multi-scale, multi-model system called Hurricane Analysis and Forecasting System (HAFS). The HAFS is NOAA's next-generation, multi-scale numerical model and data assimilation package, which will provide an operational analysis and forecast out to seven days. This will provide reliable and skillful guidance on Tropical Cyclone (TC) track and intensity (including RI), storm size, genesis, storm surge, rainfall, and tornadoes associated with TCs, all within the framework of the Unified Forecast System (UFS) and its rolling three-year Strategic Implementation Plan (SIP). Central to the development of HAFS will be the FV3 dynamical core, with embedded moving nest capable of tracking the inner core region of a hurricane at 1-2 km resolution. Section 13 discusses the future of HFIP.

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[5] (Cangialosi, June 2018). Results from HFIP model guidance are then compared with the baseline to assess progress. Fig. 1 shows the mean absolute errors of the consensus over the period 2006-2008 for the Atlantic basin. A separate set of baseline errors (not shown) was computed for the eastern North Pacific basin (Franklin, 2009, 2010).

To provide a more representative, longer-term perspective, the progress of HFIP models is 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. Fig. 1 also shows the skill 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.

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

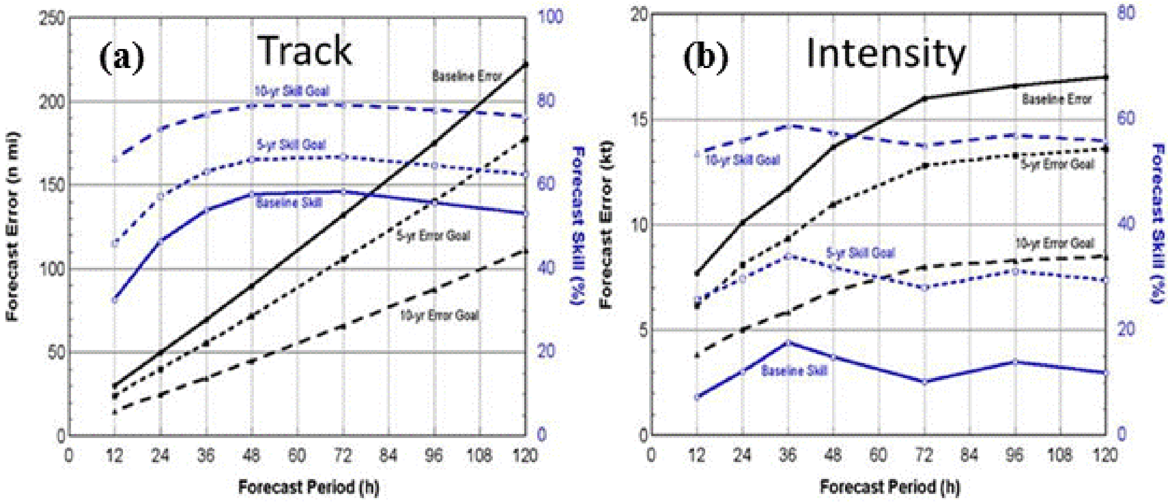

Figure 1: HFIP (a) Track and (b) Intensity Error Baseline and Goals, where the forecast errors are represented by black lines labeled on the left side of the graph, and the forecast skill is represented by blue lines labeled on the right side of the graph. Solid black lines represent baseline forecast 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 data DA techniques. Most of the GFS developments have been at 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.

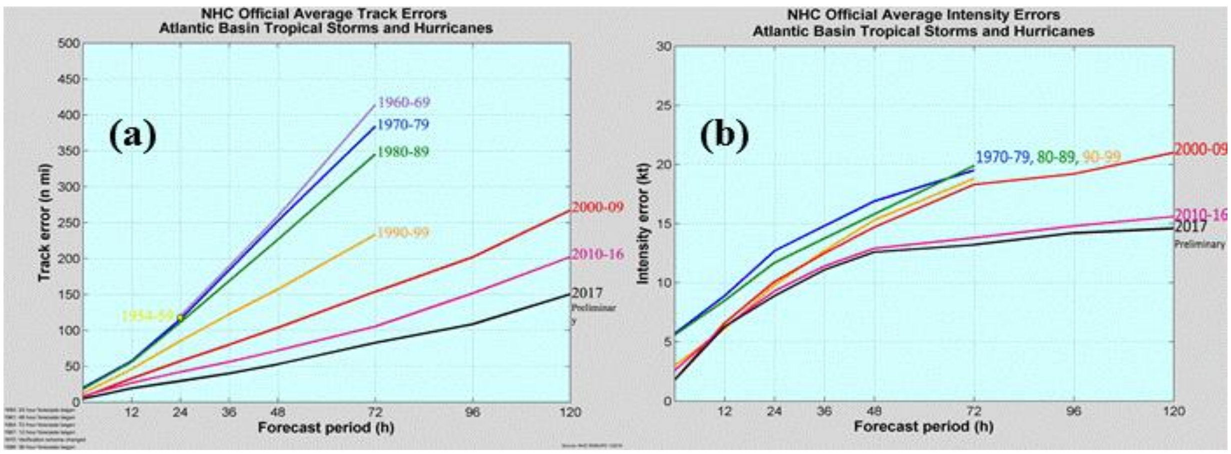

Figure 2: Official NHC (a) Track errors (1960-2017) and (b) Intensity errors (1970-2017) in the AL basin.

While significant track improvements have been achieved since the 60's, Fig. 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, the larger they are (Zhang et. al., 2016; and Alaka et. al., 2017). The Basin-Scale HWRF shown in the cover picture has demonstrated the usefulness of expanding the regional domain for TC predictions. Nevertheless, the operational TC regional models, both HWRF and HMON, are configured to be smaller than the Basin-

Scale HWRF, but larger than typical CONUS regional domains. These TC regional models are further (one-way) nested within the global models, to provide seamless track and intensity predictions.

5. Operational HWRF and HMON systems (Stream 1)

a. HWRF System

One of the major accomplishments of HFIP is 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 now one of the top-performing track prediction models, and is now 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 a part of the WRF infrastructure (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.

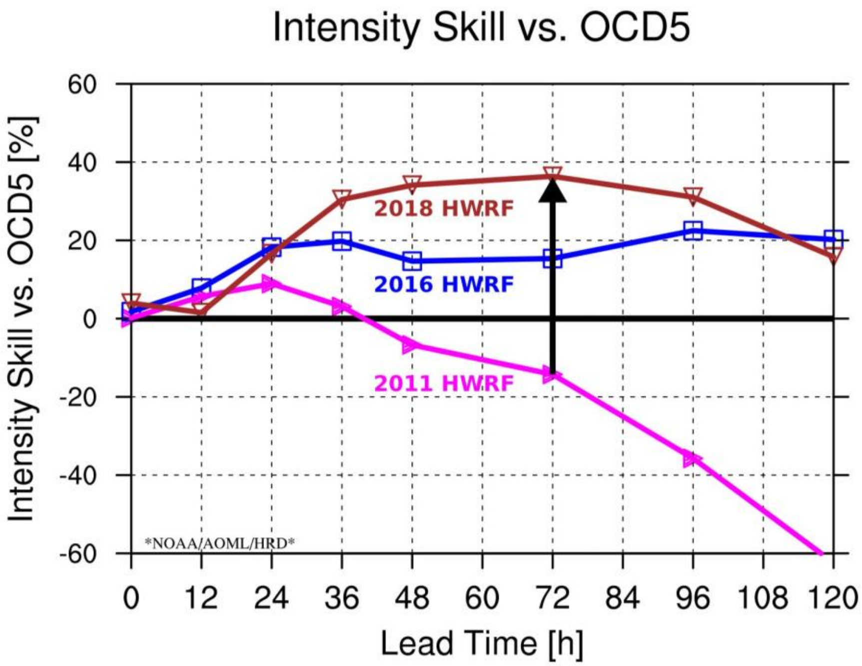

Figure 3: HWRF intensity skill relative to Decay-SHIFOR for the 2011-2018 Atlantic seasons.

Fig. 3 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 seven years (2012-2018), the HWRF system was upgraded considerably under HFIP year after year.

- In 2012, for the first time, the doubly-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.

- 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. 3).
- 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.

Clearly, steep-step progress is being made under the HFIP with every yearly upgrade. HWRF has improved by about 40-60% since 2011 (Fig. 3). Consistent with Fig. 3, HWRF is the 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. HWRF has been the most reliable intensity prediction tool in other global basins as well (Atlas et. al., 2015) (see details in section 5d).

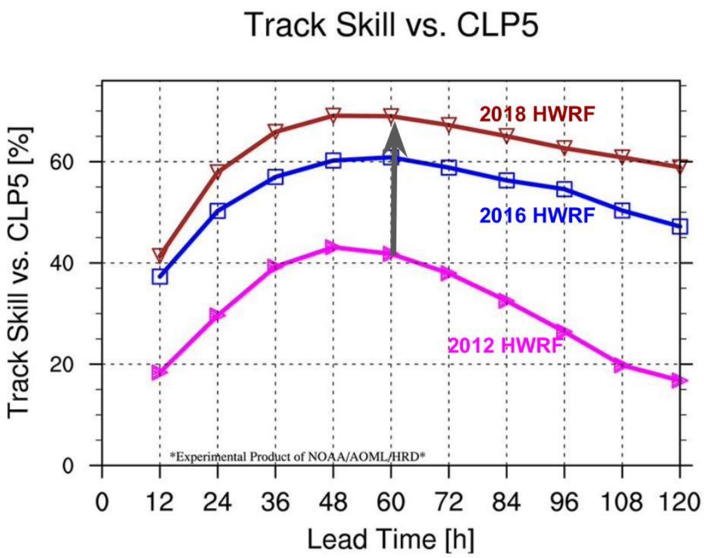

Figure 4: HWRF Track skill relative to CLIPER5 for 2011-2018 Atlantic seasons.

Fig. 4 illustrates the improvements to track forecasts from the HWRF system since 2011, as measured in terms of skill relative to CLIPER. As mentioned earlier, HWRF was initially developed for improving intensity guidance. However, because global models continue to lack the resolution to capture the inner core structure required to produce intensity forecasts, HWRF has also been used to provide some reliable track guidance, together with GFS and other models. Clearly, HWRF has improved track guidance by 30-40% since 2011. Nevertheless, it should be noted that HWRF is a regional model that uses boundary conditions from GFS. Thus, any improvements to the GFS would positively impact the HWRF system.

b. HMON System

Hurricanes in a Multi-scale Ocean-coupled Non- hydrostatic model (HMON) was developed to provide higher-resolution intensity forecast guidance to NHC, along with HWRF. HMON replaced the legacy (hydrostatic) Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model, 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 other 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 two hurricane seasons, and has demonstrated forecast consensus improvement.

c. HWRF/HMON Results from the 2018 Season

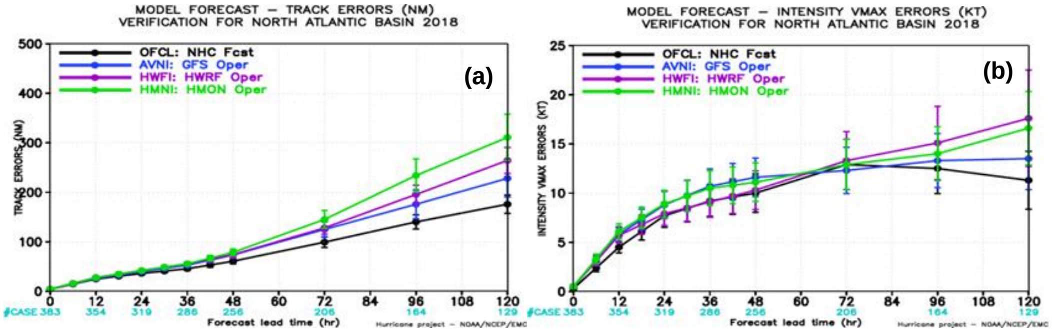

Figure 5: Verification results for (a) track and (b) intensity forecasts in the North Atlantic Basin for 2018.

For the 2018 Hurricane season, NCEP dynamical models performed well (Fig. 5). As expected, GFS was the best-performing model for track prediction, followed by HWRF and HMON (Fig. 5a). GFS was comparable to the official forecasts from NHC up to 72 hours. For intensity (Fig. 5b), HWRF had the lowest intensity errors and was comparable to the official forecasts from NHC (in black) up to 72 hours, but the skill of HWRF dropped sharply after that. Surprisingly, at longer lead times, GFS was the most skillful in terms of mean intensity error as well.

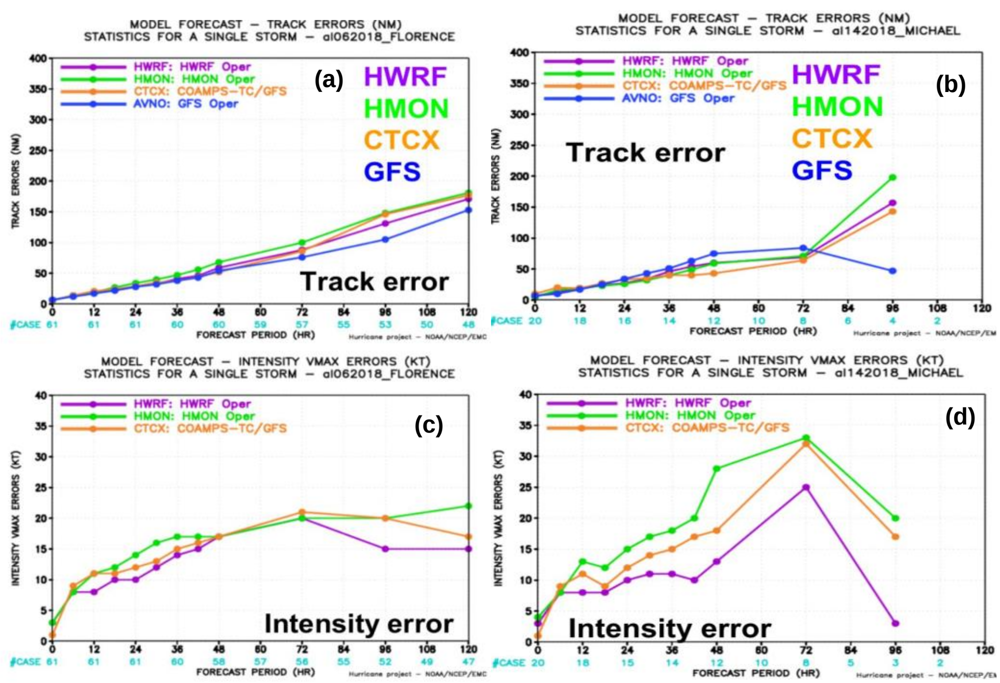

Figure 6: 5-day Track (top) and Intensity (bottom) forecast verification for (left) Hurricane Florence and (right) and Hurricane Michael.

HWRF performed well for both of the major landfalling hurricanes in the Atlantic basin, namely, Florence and Michael (Fig. 6). GFS performed the best for tracks in either storms. HWRF and HMON were very similar to GFS in track error up to 72 hours. HWRF had the best intensity forecast performance, even outperforming NHC official forecasts for some of the forecast intervals (not shown). Some cycles of HWRF forecasts captured the RI of Hurricane Michael at least 4 days in advance (Fig. 6d). The major intensity errors from Michael were mostly associated with early landfall, consistent with the small but noticeable tracks errors in Fig. 6.

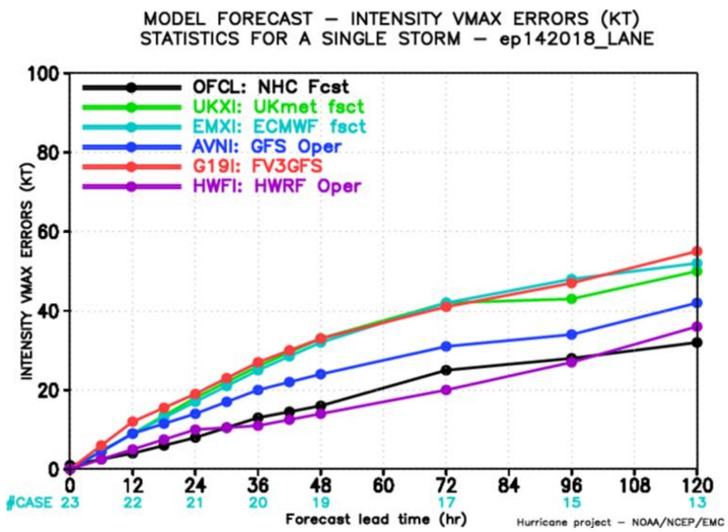

Figure 7: 5-day Intensity forecast verification for Hurricane Lane.

For the first time, during hurricane Lane, a P3 aircraft was flown over the Central Pacific for assimilating inner-core winds into the HWRF model. Both HWRF and GFS performed well for Hurricane Lane in the East Pacific basin (Fig. 7). Intensity forecast performance for Hurricane Lane was again the best for operational HWRF, showing lower errors compared to NHC official forecasts for most of the lead times.

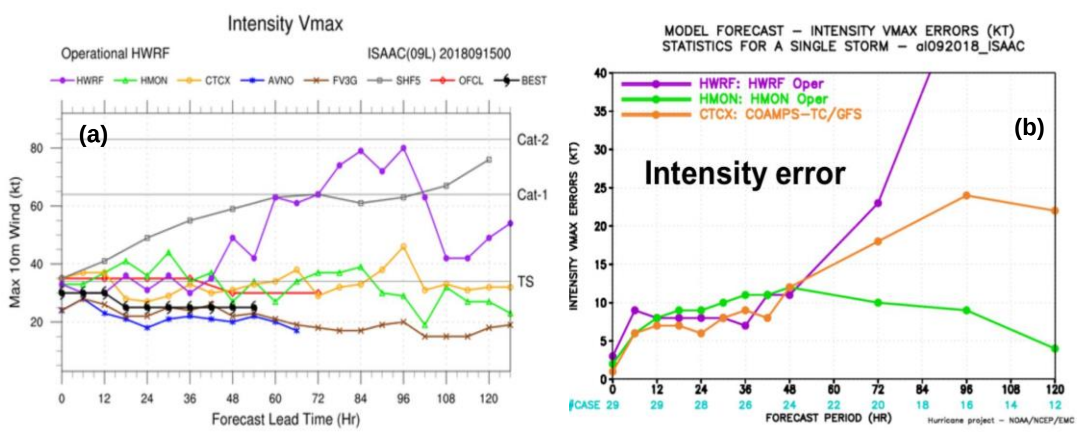

Figure 8: (a) Max 10 m wind from different models, for one of the cycles from Isaac (2018091500) and (b) intensity forecast verifications for TC Isaac.

The most challenging outlier event for HWRF in the 2018 season was TC Isaac. While Isaac rapidly weakened as a tropical storm after crossing the Lesser Antilles, HWRF continued to strengthen the storm to a major hurricane (Fig. 8a). This led to large intensity errors beyond 48 hours (Fig. 8b). Post-analysis of the season illustrated that almost all the larger errors beyond 72 hours, when compared to the official forecasts, may be attributed to this false alarm from HWRF (Section 7c). The basin scale HWRF, which is the same version of HWRF but for larger domain and multiple moving nest; and capable of tracking any number of TCs in the domain, was continuously running in parallel under Stream 2 (Section 8a). The results from basin scale HWRF showed superior skills for Isaac intensity forecast, and was as good as the operational HWRF for all other Atlantic hurricanes in 2018. Post-analysis showed that the multiple moving nest, which provided higher resolution, not only around Isaac but also nearby storms, Hurricanes Florence and Helene, captured storm-storm interactions better, illustrating the need for hurricane moving nests in Next Generation Global Prediction System (NGGPS).

d. HWRF Performance in other Global Basins

During the 2018 season in the western North Pacific basin, the Joint Typhoon Warning Center (JTWC) official track forecast proved to be the most skillful for virtually all forecast periods (Fig. 9a). Among the guidance models, HWRF and GFS continued to be top performers. The COAMPS-TC/GFS track skill was competitive with the HWRF and GFS from 12-48 h, after which it began to trail the other models.

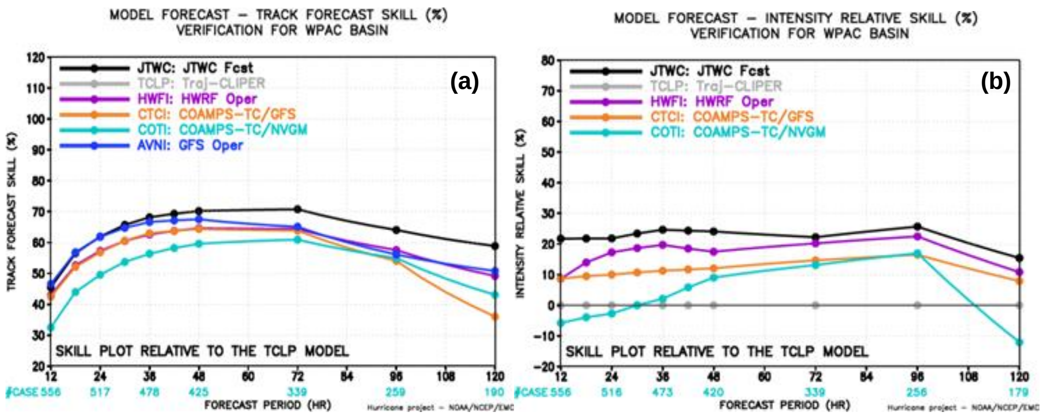

Figure 9: Western North Pacific (a) track forecast skill, and (b) intensity forecast skill.

For intensity in the western North Pacific (Fig. 9b), the JTWC produced the most skillful forecasts for nearly all forecast cycles, and these forecasts were especially skillful at the shortest and longest lead times. Operational HWRF came closest to JTWC in terms of intensity relative skill, and was consistently better than COAMPS-TC/NVGM and COAMPS-TC/GFS at all lead times.

e. Prediction of Rapid Intensification

Predicting the RI of TCs is a complex, challenging, and important forecast problem. In general, apart from the well-documented impacts of the upper ocean on intensity changes, environmental factors such as wind shear, moisture in the low to mid troposphere, and inner-core processes - ranging from convective to mesoscale - all have been known to influence the RI of TCs. All these factors interact in a nonlinear fashion, making the RI problem a complex forecasting challenge (Chen and Gopalakrishnan, 2015).

RI in hurricanes is defined as an increase in sustained 1-minute, 10 m wind speed of $\geq$30 knots in a 24-hour period. In order to understand how model wind distributions correspond to reality, probability distribution functions (PDFs) for HWRF and HMON intensity changes were compared to best-track data in all three basins. Fig. 10 (top row) demonstrates that the PDFs for 24-h intensity change are similar among Best Track (black), HWRF (magenta), and HMON (green). The mean intensity change of all three is very similar. It should be noted that RI and RW lies in the 95th and 5th percentile, respectively of the PDF. High resolution models like HWRF and HMON may be able to reproduce intensity changes at these extreme ends of the spectrum. The model climatology is close to the Best Track estimates. Stratifying the mean intensity error statistics across all three basins, by including all 2018 TCs that experienced at least one RI event (Fig. 10d, 10e & 10f), makes clear that HWRF produced the lowest intensity error for such TCs.

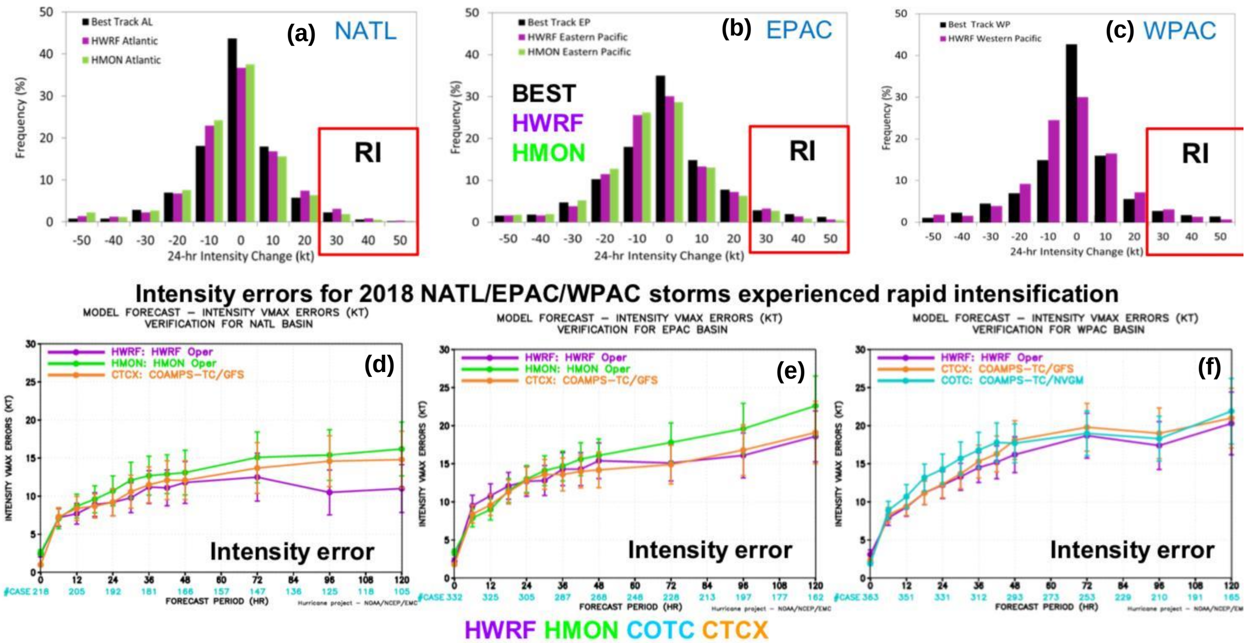

Figure 10: Top row (a, b, c): 24-h intensity change for HWRF, HMON, and the Best Track (BEST). Bottom Row (d, e, f): Intensity errors from multiple forecast cycles that had at least one RI period.

Indeed, consistent with the earlier findings from the PDF of 24-h intensity change (Fig. 10a, 10b, 10c), in 2018, some forecast cycles of HWRF captured the RI of Hurricane Michael at least 4 days in advance (Fig. 11a). It should be noted that this storm developed in a hostile environment of shear exceeding 20-25 knots, where RI predictions may be a challenge. However, while HWRF captured RI fairly well for the one complex case of Michael, RI of TCs continues to pose the greatest challenge for forecasting (Section 7c). In 2018, for instance, there were several short-lived RI events during Hurricanes Alberto, Beryl, Chris, and Oscar (Fig. 11b). All of them were missed by HWRF. Section 7c discusses the challenges with operational RI prediction in more detail.

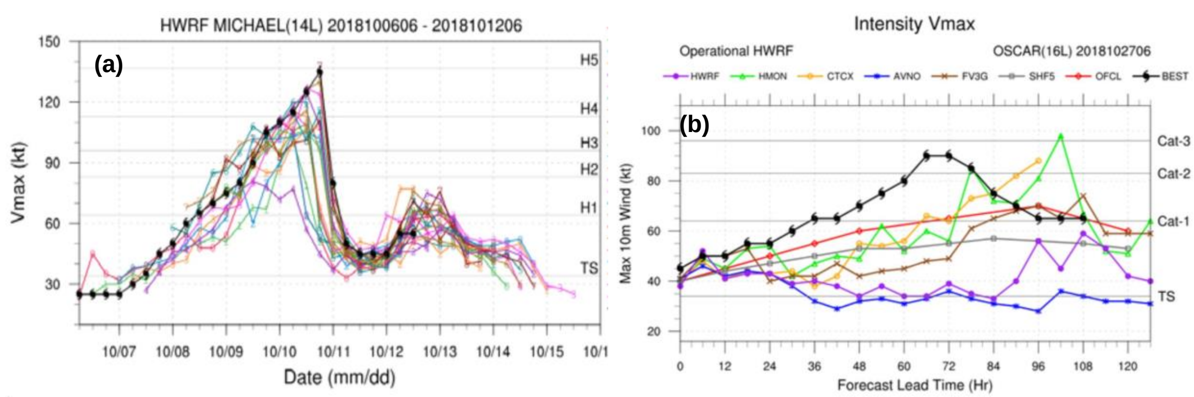

Figure 11: (a) HWRF intensity predictions from various forecast cycles for Hurricane Michael, compared to Best Track estimates. (b) A cycle of HWRF along with other model forecasts, for Hurricane Oscar, compared to Best Track estimates.

6. The HWRF Data Assimilation (DA) system

One of the major accompaniments of HFIP was the development of the state-of-the-art inner-core DA system for HWRF in 2017. In an earlier report, we gave an overview on the DA systems used in HFIP[6]. As in 2017, the 2018 HWRF used a fully-cycled hybrid DA system. The covariance for the inner nest of this system is provided by a cycling Ensemble Kalman Filter (EnKF), which can represent inner core structures much better than the covariance provided by Global Data Assimilation System (GDAS). The workflow for this system is illustrated in Fig. 12. Note that for operational reasons, there are only sufficient resources to run this system for one storm at any given time, and that other non-priority storms must continue to use GDAS covariance for inner nest DA.

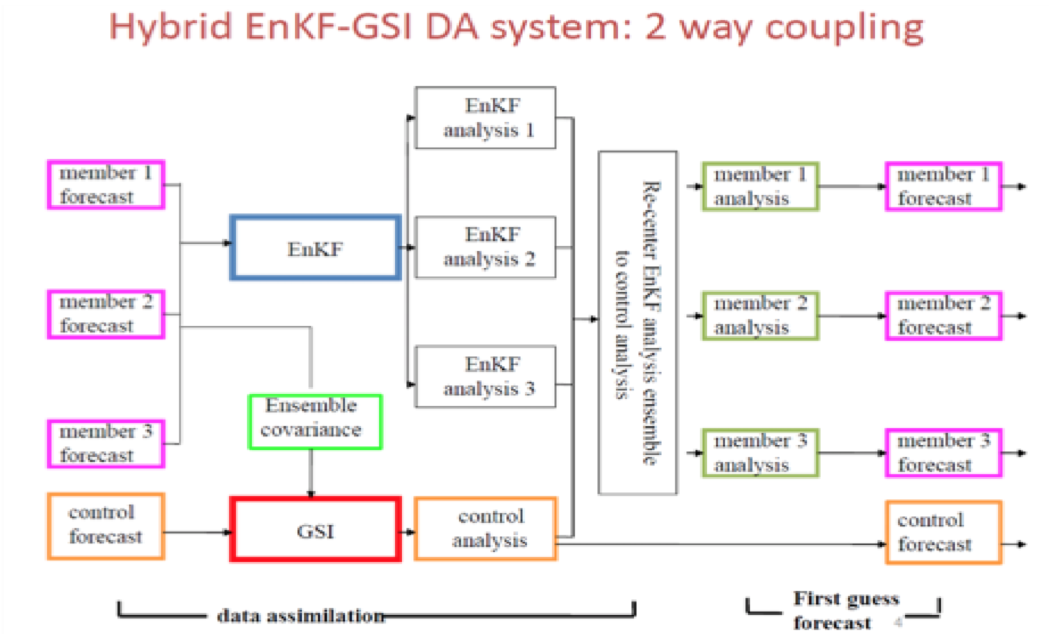

Figure 12: Diagram illustrating the new HWRF DA system, with full covariance cycling provided by EnKF.

One significant upgrade to the 2018 HWRF DA system was the addition of stochastic physics to the HWRF DA ensemble perturbations that are used to formulate mesoscale covariance. The Cumulus and the PBL parameterizations schemes were perturbed at each call to physics. In particular, perturbations were applied to (a) the convective trigger function within the Simplified Arakawa Schubert (SAS) cumulus parameterization scheme, (b) the planetary boundary layer (PBL) height within the Global Forecast System (GFS) PBL scheme, and (c) the drag coefficient (C_D) within the modified Geophysical Fluid Dynamics Laboratory (GFDL) surface-layer scheme. This variability leads to greater ensemble spread, ultimately drawing the ensemble analysis more closely to the observations. The result of this is superior performance, in terms of both track and intensity (Fig. 13).

---

[6] http://www.hfip.org/documents/HFIP_AnnualReport_FY2016.pdf

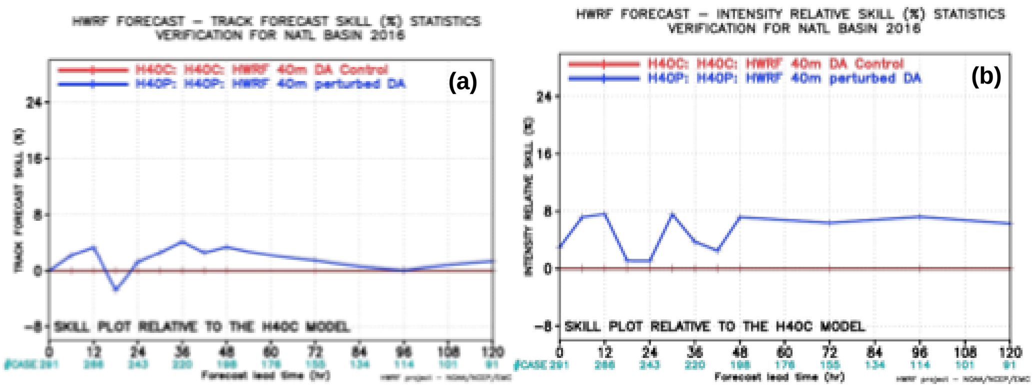

Figure 13: Impact on HWRF track (left) and Vmax (right) skill, from stochastic DA ensemble perturbations, as compared to a baseline without perturbed physics.

A number of additional datasets were also added to the 2018 HWRF DA system. Among these upgrades was assimilating radial velocity from the NOAA Gulf IV (GIV) tail Doppler radar, in addition to surface wind speeds derived from the Stepped-Frequency Microwave Radiometer (SFMR) aboard WP-3D aircraft. The impact of the SFMR data, in particular, was found to considerably improve the short-term track forecast (see Fig. 14).

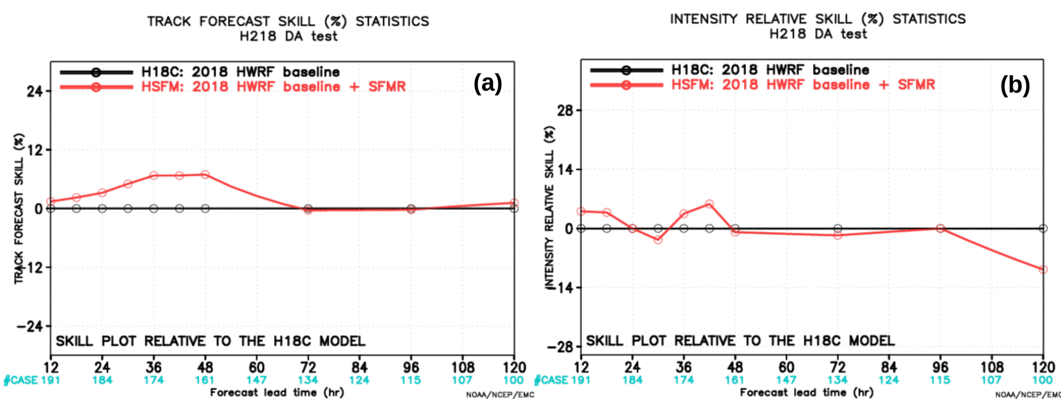

Figure 14: Impact on HWRF skill in predicting track (left) and maximum wind speed (Vmax, right), due to assimilation of SFMR wind-speed observations.

In addition to these changes, a significant change was made to how HWRF uses dropsonde data. Operationally, dropsonde information is sent in the World Meteorological Organization (WMO) TEMPDROP format, which only transmits the release location with the data. This presents a major problem in the high-wind regions of tropical cyclones, where significant lateral advection of the dropsonde takes place. Some dropsondes have been observed to travel 180 degrees around the eyewall during their transit, and if sondes with such large displacement are assimilated as a single vertical column, severe negative consequences arise during data assimilation. Historically, NCEP has handled this problem by disregarding dropsonde wind observations in high-wind regions.

Recently, code was added to HWRF that allows the lateral advection of dropsondes to be calculated from data contained within the WMO TEMPDROP message. With this change, the dropsonde location can be estimated to within roughly 1 km, which is less than current model grid spacing. Thus, it is no longer necessary to disregard any dropsonde wind observations. Testing has shown that accounting for drift improves the intensity forecast by 10-15% before 72h (Fig. 15). There is a relatively minor impact on the track forecast, though it does improve more often than not (not shown).

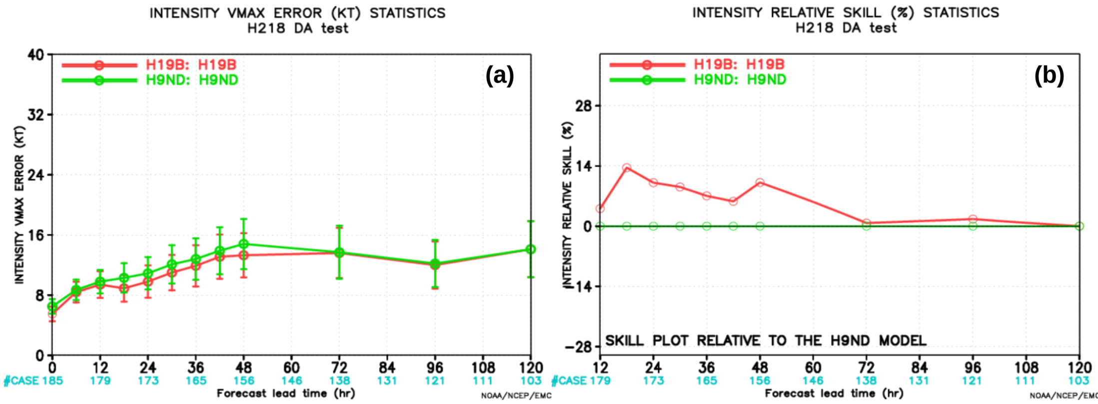

Figure 15: Impact on HWRF Vmax error (left) and skill (right) of considering dropsonde drift in HWRF.

Improvements in DA, particularly for the inner core, mean that inner-core reconnaissance is becoming increasingly important for improving intensity forecasts. HWRF now uses all reconnaissance data that is transmitted in the operational data stream, and is the only operational model in the world to do so.

7. Operational Hurricane Guidance Improvements

The 2018 Atlantic hurricane season activity was above normal with 15 named tropical cyclones, eight of which became hurricanes. Two of the hurricanes became major hurricanes (Florence and Michael). Using the best track data as of early March, 2019, there were 20 occurrences of RI using the definition of a 24-h period with a 30 kt or greater increase in the maximum wind (30kt/24hr), corresponding to about 6% of the 2018 sample 24-h intensity changes. This percentage is very close to the long-term mean. The majority (15) of the RI cases were from Hurricanes Florence and Michael. The other five were from Hurricanes Beryl (1), Chris (2), Oscar (1) and Tropical Storm Nadine (1).

NHC uses several deterministic guidance models for their official intensity forecasts, including NCEP's HWRF and HMON and the Navy's COAMPS-TC 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, COAMPS-TC regional models. IVCN runs when two or more of the above models were available. In 2018, COAMPS-TC was only available early enough for the interpolator to run for about 70% of the forecast cases, so it will not be included in the RI verification results below. However, the IVCN results do include COAMPS-TC for the times when it was available. This problem should be resolved for the 2019 hurricane season because COAMPS-TC will be transitioned from research to operations by the Navy.

a. Track Guidance

In 2018, official Atlantic track forecasts (Fig. 16a) were very skillful, and close to or better than the best-performing consensus aids - HCCA and Track Variable Consensus Approach (TVCA) (Cangiolosi, 2018). EMXI was the best dynamical model at 48h and beyond, and had nearly the same skill as the consensus aids at those lead times. GFSI was the best dynamical model at 24h and 36h, but not as skillful after that. HWFI, AEMI, and EGRI were fair performers. HMNI was competitive early, but its skill dropped off after 48h. Performing less competitively were NVGI and CMCI.

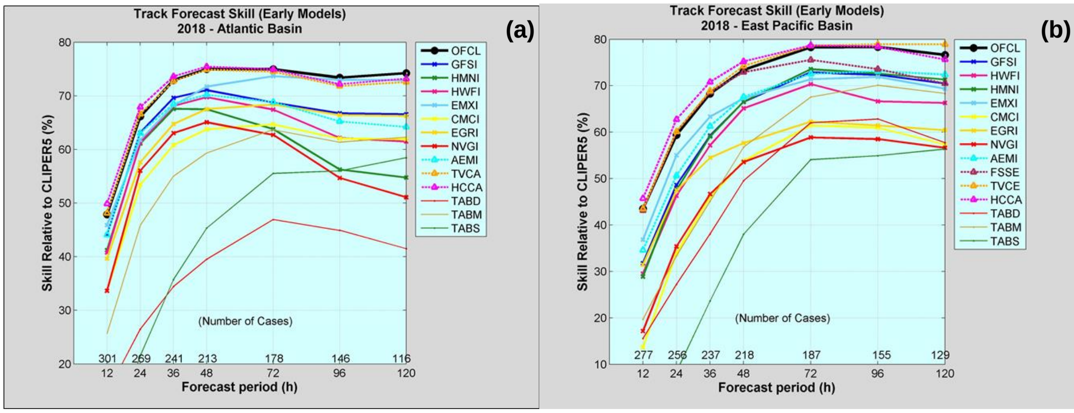

Figure 16: Official track forecast skill in 2018 for the (a) Atlantic (left) and (b) eastern Pacific (right) basins. Numbers immediately above the X-axis show the total number of cases covered by each data point.

In the eastern Pacific (Fig.16b), the consensus aids HCCA and TVCE led the way with the highest skill. NHC official forecasts were very skillful, with performance close to that of these best consensus aids. FSSE had skill close to the best consensus aids in the short term, but trailed slightly at longer leads. EMXI was the best individual model through 36 h, while GFSI, AEMI, and HMNI did better than EMXI at leads longer than that. HWFI was just behind the best individual models; EGRI started well, but trailed beyond 36h. CMCI and NVGI were not competitive.

b. Intensity Guidance

Intensity forecast verifications for the 2018 season are shown in Fig. 17. In the Atlantic basin (Fig. 17a), official forecasts were better than all of the guidance at lead times of 12h, 96h, and 120h. IVCN and HCCA were the best consensus aids, with IVCN best from 12 h to 48 h, and HCCA best at 120 h. HWFI was the best individual model through 48 h, with skill comparable to the official forecast for leads of 24 to 72h, while GFSI was the best individual model at lead times of 72 h and longer. HMNI was not as skillful as HWFI or GFSI, but was more skillful than the statistical aids. DSHP and LGEM were skillful, but had less skill than dynamical models at most time period. EMXI showed some skill but was not as competitive.

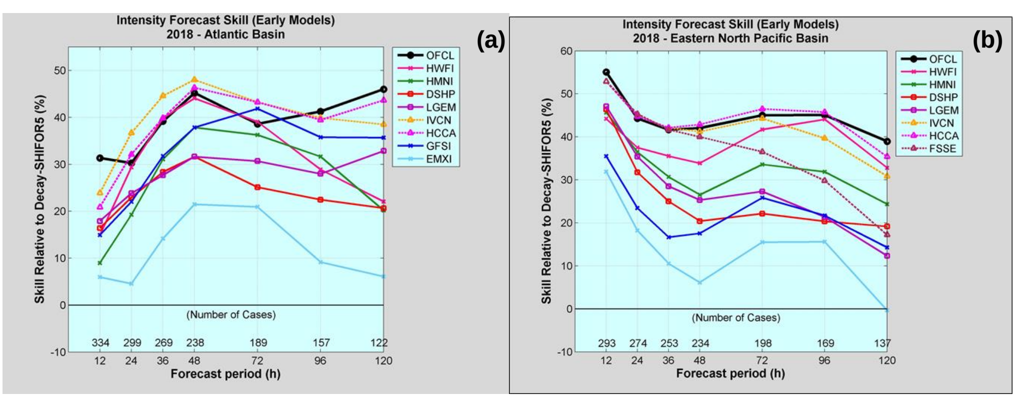

Figure 17: Official intensity forecast skill in 2018 for the (a) AL (left) and (b) EP (right) basins.

In the eastern Pacific (Fig. 17b), official intensity forecast performance was better than or comparable to the best consensus aids (IVCN and HCCA). FSSE was one of the best aids early, but it trailed off after 48h forecast period. HWFI was the best individual model beyond 24 h. Though HMNI was not as skillful as HWFI, it did beat the statistical aids. DSHP and LGEM were moderately skillful, but were not competitive with the best aids.

c. Rapid Intensification/Weakening Prediction

The operational intensity verification described in section 7b showed basin-wide statistics. A new metric being proposed to track HFIP progress is the IVCN intensity error for Atlantic cases with at least one RI occurence in any 24 hr interval during the forecast period, relative to the 2015-2017 baseline. Fig. 18 shows the IVCN mean absolute error (MAE) for the RI, no-RI and total samples for 2018 and the baseline period. Note that there are no RI cases for the 12-h forecasts because RI is defined in terms of a 24-h intervals. Also note that the RI sample size can increase with forecast length even though the total sample decreases with forecast length because the RI sample only requires one 24 h period with an RI event, which can occur in multiple 24 h intervals after 24 h.

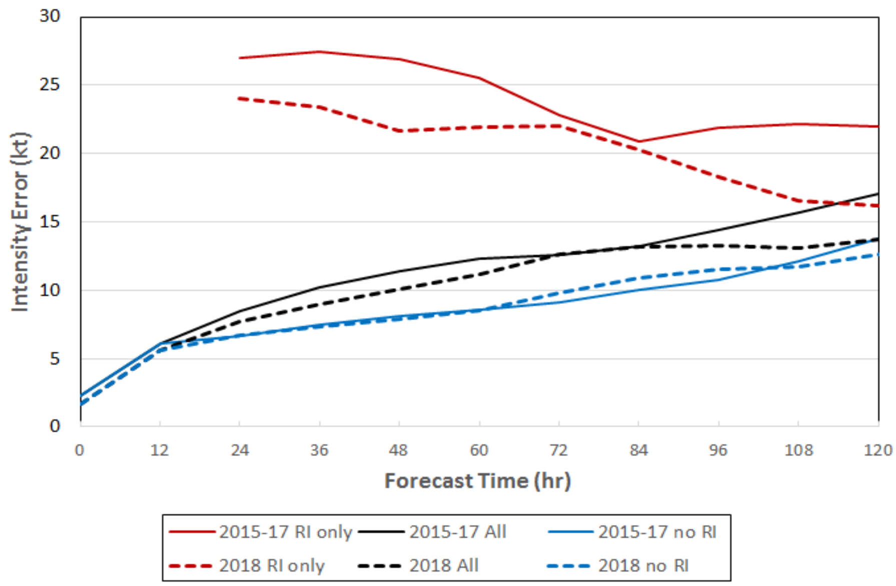

Figure 18: The mean absolute intensity error for the RI cases, non-RI cases, and total sample for 2018 and the 2015-2017 baseline. The sample sizes at 24, 48, 72, 96 and 120 h for the 2018 RI cases are 20, 42, 49, 48, and 46, respectively; for the 2015-2017 baseline, sample sizes are 74, 119, 138, 134, and 136, respectively.

Fig. 18 shows that the errors for the RI cases for 2015-2017 and 2018 are much larger than those for the non-RI cases, especially at the earlier times. These errors are much larger than the typical uncertainty in the intensity estimates, making it easier to track progress than if the total sample errors were used. Comparing the 2018 errors to the 2015-2017 baseline for the non-RI cases shows there is little difference at all forecast periods. However, for the RI cases, the 2018 errors are smaller than those for 2015-2017 suggesting some progress has been made. However, the 2018 RI sample size is fairly small (maximum of 49 cases at 72 h), so it will likely take several more seasons to confirm this result.

Table 1. The metrics from the 2x2 contingency table used to evaluate the RI forecasts.

Metric	**Formula from Contingency Table elements**
Probability of Detection (POD)	a/(a+c)
False Alarm Ratio (FAR)	b/(a+b)
Threat Score (TS)	a/(a+b+c)

The ability of a model to predict RI can be considered as a classification problem. Again using the 30kt/24 hr RI definition, the standard 2x2 contingency table can be populated from the verification sample, where the four elements of the table are as follows: The number of times when a) RI was predicted and observed; b) RI was predicted but not observed; c) RI was not predicted but was observed; d) RI was not predicted and not observed. Several metrics can be calculated from the contingency table.

Because RI is a rare event, the largest fraction of cases will be in element (d) of the table, which gives misleading results when that appears in the denominator. Thus, three metrics were considered that only use elements (a), (b) and (c) as shown in Table 1.

The POD is a measure of how often the model correctly forecasted RI, the FAR is the fraction of times where the model forecasted RI but it did not occur (over prediction). Threat Score (TS) can be interpreted as the ratio of the area overlap between the forecasted and observed cases, to the total area of the observed and forecast cases. It can be shown that a random forecast of RI based on the climatological RI probability P will have a TS of P/(2-P). The long-term RI probability for the Atlantic sample is about 6%, which would give a TS value of 0.031 (3.1%). Thus, a TS greater than this value is considered to have skill.

The POD and FAR were evaluated for the NHC Official (OFCL) forecast and all of the models included in IVCN except COAMPS-TC because of the small sample size for that model. The POD and FAR were evaluated over the intervals 0-24, 24-48, 48-72, 72-96, and 96-120 h. This is a fairly stringent evaluation because the forecast has to have the timing of the RI event exactly correct to obtain a positive contribution to the POD.

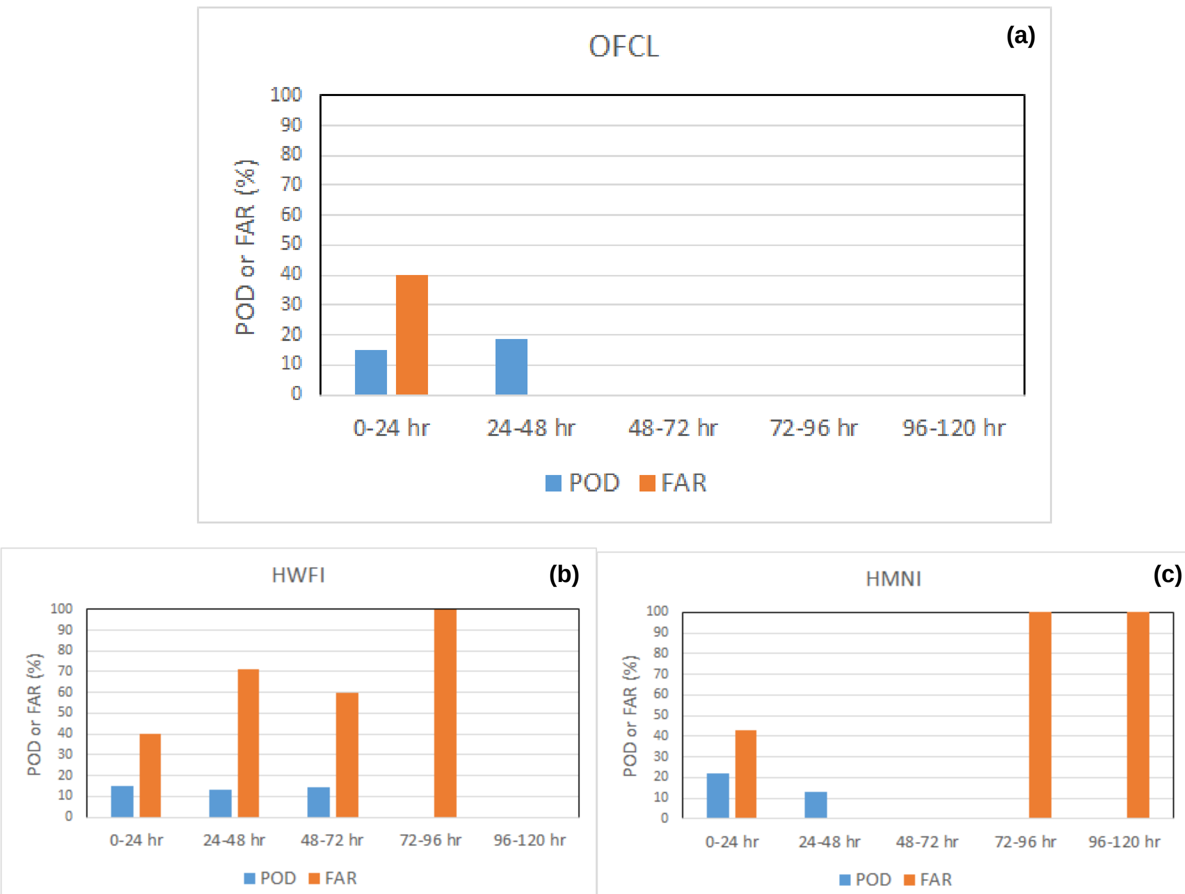

Figure 19: The probability of detection (POD) and False Alarm Rate (FAR) of RI forecasts from the 2018 Atlantic sample for the NHC Official (OFCL) and early versions of the HWRF (HWFI) and HMON (HMNI) models.

Fig. 19 shows the POD and FAR for OFCL, HWFI and HMNI. The LGEM and DSHP statistical models did not forecast any RI events, and IVCN only forecasted one event for the 0-24 hr period, so those models are not included in the figure. The NHC Official and HWRF and HMON models correctly forecasted RI for 10 to 20% of the cases through the 48-72 hr forecast interval. HWRF and HMON correctly predicted the RI for two of the eight Florence cases, but missed the other six. HMON also captured one of the seven Michael RI events at the 0-24 h interval. For nearly all time intervals, FAR is much larger than the POD indicating that when the models did predict an RI event, it usually was not observed. These results suggest that neither NHC nor the regional models were very successful with predicting RI for the 2018 Atlantic cases, but did pick up a few cases for Florence and Michael.

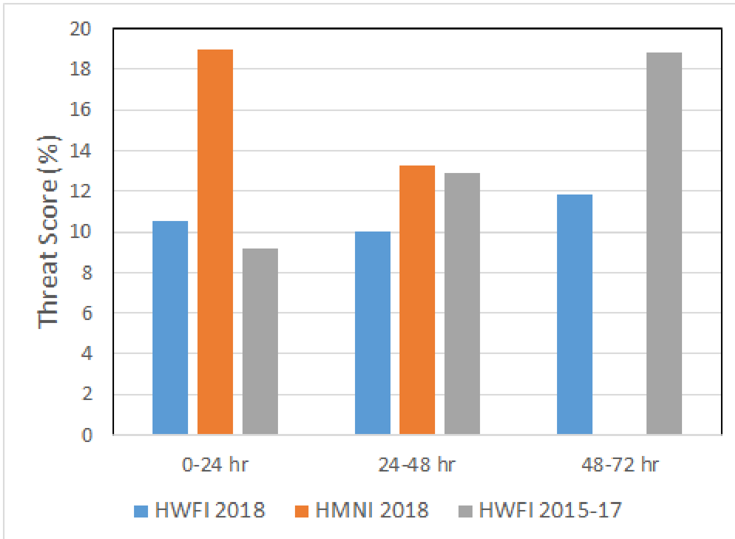

Figure 20: Threat Score (%) for the early versions of the HWRF and HMON models for the 2018 Atlantic basin. The HWRF values are also shown for the baseline period (2015-2017).

Fig. 20 shows the Threat Scores for both HWRF and HMON for 2018, and for HWRF during the baseline period. HMON was not available before 2017, so its baseline TS could not be calculated. The results are shown only through the 48-72 h interval since the TS values are all zero for the longer intervals. This figure shows that the TS values ranged between 9 and 19% and the 2018 values were comparable to the HWRF values for the baseline period. As described above, a random RI forecast based on the climatological RI probability would give a TS of 3.1%. The values for HWRF and HMON exceed that value through the 48-72 hr indicating that the models did have some relative to a climatological forecast.

Because of the fairly low POD by the deterministic intensity models, NHC also uses several statistical post-processing techniques to estimate the probability of RI based on predictors from the TC environment (using GFS forecast fields) and models forecasts. In 2018, NHC ran four statistically-based RI probability models. Three of these uses a subset of the predictors from the SHIPS model as input to a discriminant analysis (RIOD), logistic regression (RIOL) and Bayesian (RIOB) techniques. The fourth RI model was developed under partial HFIP support, and uses the forecasts from several deterministic intensity models as input to a logistic regression technique. This model is called the Deterministic to Probabilistic Statistical (DTOPS) model, and was run in real time on WCOSS for the first time in 2018. All four of these models estimate the probability of RI in terms of the traditional 30 kt or more increase in 24 hr and six other increases over various time intervals. The verification included three definitions including 30 kt or more in 24 hr, 55 kt or more in 48 h (55kt/48hr) and 65 kt or more in 72 hr. (65kt/72hr).

A common metric for evaluating probabilistic forecast is the Brier Score (BS), which is the sum of the squares of the difference between the forecasted probability and verifying probability (either 0 or 100% for non-RI and RI cases, respectively), divided by the sample size. BS is then converted to a Brier Skill Score (BSS) by calculating the percent reduction in the BS, relative to a forecast where the probability is set to the long-term climatological value. Positive BSS indicates forecast skill. For comparison, the BSS can also be calculated for the deterministic models by assigning a probability of 100% if the model forecasted RI and 0% if it did not.

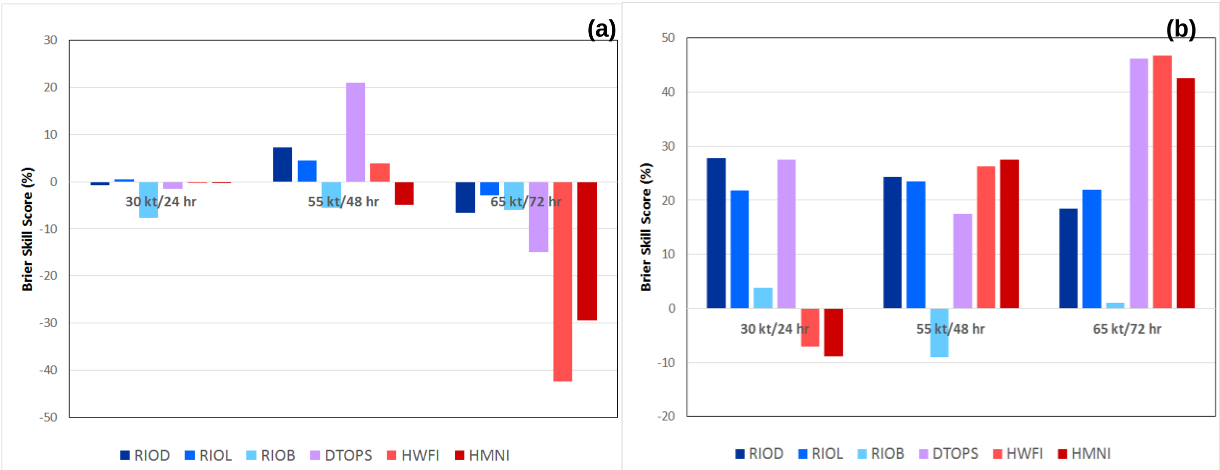

Figure 21: Brier Skill Score from four statistical post-processing probabilistic RI models and the early HWRF and HMON models converted to a binary (100% or 0%) probabilistic forecast for (a) the 2018 Atlantic sample and (b) eastern and central North Pacific samples.

Fig. 21a shows the BSS for the four statistical RI models, HWRF and HMON. This figure shows that none of the models had skill for 30kt/24h or 65 kt/72h but some of the statistical methods and HWRF had some skill for 55 kt/48 h. The best performer was DTOPS for the 55kt/48 hr threshold. This figure shows that the statistical methods out-perform the dynamical models for RI prediction, even though the dynamical generally have smaller average errors for the total sample as was shown in Section 7b.

The BSS values in Fig. 21a were much smaller than in previous years, especially for 30kt/24hr and 65kt/72hr, which typically range between 10 and 20%. This lack of skill suggests that the 2018 Atlantic RI forecasts were more difficult than usual. To get an expanded sample for evaluation of the 2018 intensity models for RI cases, the eastern (EP) and central (CP) North Pacific forecasts were also verified. The EP basin had above normal TC activity, with 22 named storms, 12 of which became hurricanes. Nine of the hurricanes became major hurricanes. Only one TC originated in the CP, but it became a major hurricane. There were 62 RI events (30kt/24hr definition) in the 2018 combined Pacific sample, which is

For the 2018 combined Pacific sample, POD, FAR, and TS were calculated for each of the NHC Official forecasts and the “early” versions of the HWRF and HMON models. The NHC Official forecasts for the 0-24 hr interval were much better for the Pacific sample than for the Atlantic. The POD for the 0-24 hr interval was 48% and was greater than the FAR value of 32%. For the later intervals, the NHC official forecast results were about the same as for the Atlantic sample. The HWRF results for the Pacific sample showed some improvement over the Atlantic, with a POD as high as 26% for the 0-24 h interval. However, the FAR values were still higher than the POD values at all forecast intervals. The HMON results for the Pacific were comparable to those for the Atlantic.

As described above, the evaluation of the RI forecasts using POD and FAR for fixed 24 h intervals does not allow for timing errors. For example, if a model predicted RI to occur but was off by 6 h in the timing, that would not count towards POD but would count towards FAR. The use of the RI definitions with the longer time intervals (55kt/48hr and 65kt/72hr) do allow for some phase errors. Fig. 21b shows the verification of the three RI definitions in terms of BSS for the Pacific sample. This figure shows that the statistical models generally outperformed HWRF and HMON for 30kt/24h, had comparable skill for the 55kt/48h and had greater skill for 65kt/72h, with BSS values as high as 47% for HWRF. All of the models except RIOB had much greater skill than for the Atlantic samples shown in Fig. 21. These results show that both HWRF and HMON had considerable skill for the RI definitions over longer time intervals, suggestions that progress is being made for RI forecasting.

8. Important Stream-2 Results

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

Although the operational HWRF system is improving intensity forecasting skill, it should be noted that it is currently configured with only one set of high-resolution nests (i.e., 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. Finally, the current operational configuration poses many challenges for advancing data assimilation techniques and for downstream applications at and after landfall. 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 tropical cyclone forecasts at high resolution, and also serves as a prototype for the development of multiple moving, multi-level nests within the global model.

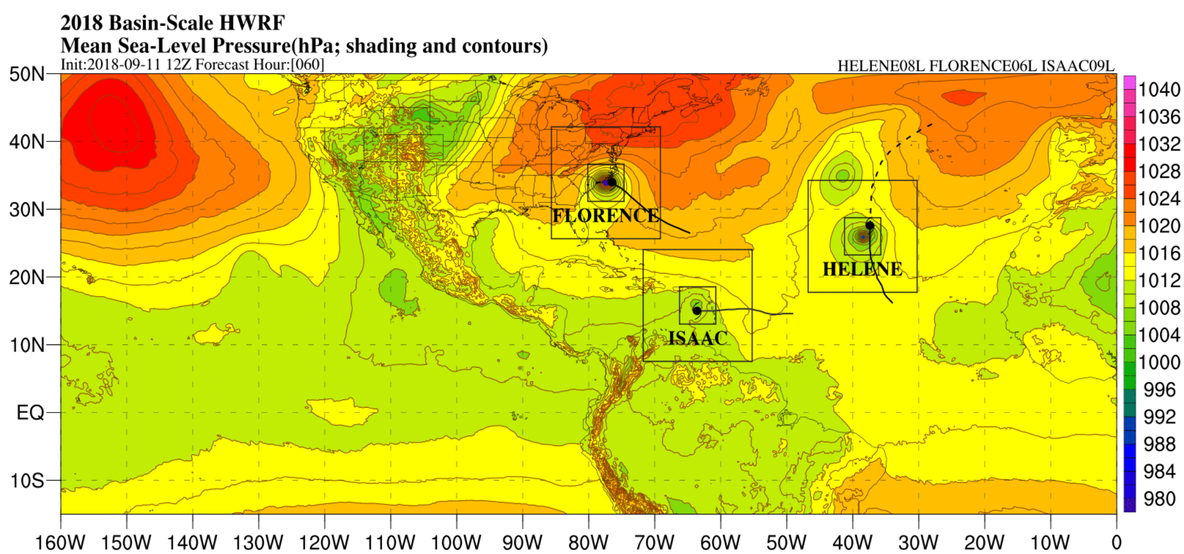

Figure 22: HWRF-B configured with three high-resolution multi-level movable nests following Hurricanes Florence, Issac and Helene for a forecast initialized at 1200 UTC 11 September 2018.

As in 2017, the experimental HWRF-B model was again run as an HFIP real-time demonstration, in parallel with operational hurricane models, during the 2018 North Atlantic and eastern North Pacific hurricane seasons. In addition to assimilating satellite, ground-based, and aircraft observations for initialization, the HWRF-B experimental system was also coupled with an advanced ocean model, and tracked up to three TCs at a time at high resolution (~1.5 km). With 2018 being a very active season in both basins, HWRF-B was able to produce forecasts for multiple TCs within the same integration. For example, nests tracked Hurricane Florence, Hurricane Helene, and Hurricane Isaac for a forecast initialized at 1200 UTC 11 September 2018 (Fig. 22). HWRF-B was at least 20-30% better than the operational HWRF for intensity predictions of these three hurricanes at longer forecast lead times. HWRF-B continued to outperform the operational HWRF, especially during the peak of the season when many TCs were active in the North Atlantic and eastern North Pacific basins. Seasonal error statistics show that the 2018 HWRF-B (HB18) produced lower track and intensity errors than the operational HWRF (H218) at longer lead times (Fig. 23). Further analysis demonstrated that capturing fine-scale details of all Atlantic storms 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 Isaac). The HWRF-B moving nests are foundational to the next generation HAFS. The advanced and well-evaluated nesting technique will be transitioned to the FV3 unified forecasting system (FV3-UFS).

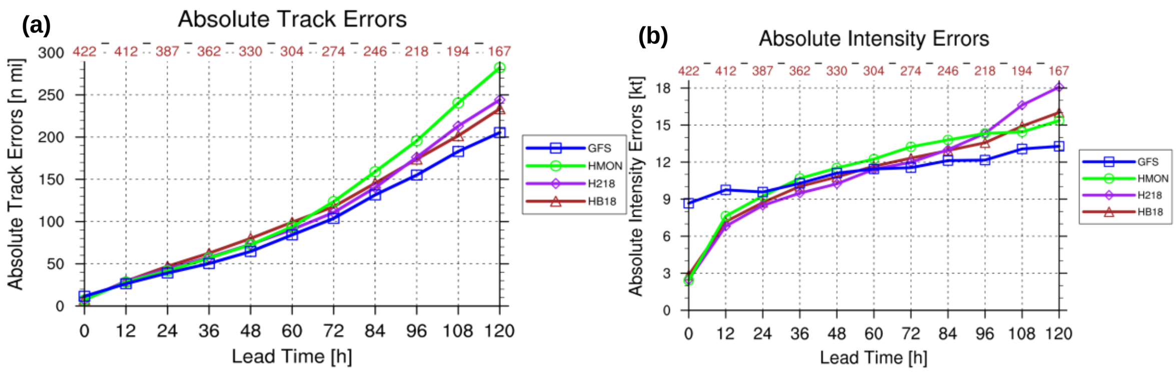

Figure 23: Verification of a) mean absolute track errors and b) mean absolute intensity errors for the 2018 HWRF-B (HB18), the 2018 operational HWRF (H218), HMON, and GFS.

b. Unified Modeling System for Hurricane Forecasting (Stream-2)

In late 2016, GFDL's Finite Volume Cubed Sphere (FV3) model was selected as the next dynamical core for NOAA's Global Forecasting System[7]. FV3 is a fully non-hydrostatic model which will replace the operational Global Spectral Core at NCEP later in 2019, with horizontal resolution of about 13 km in the upcoming implementation.

In light of the NWS's commitment toward a unified modeling approach in the next generation hurricane prediction system, one of the powerful benefits of the FV3 system is its capability to utilize two kinds of downscaling techniques for higher horizontal grid resolution, (i) Grid stretching and (ii) Telescopic nesting (Harris et. al., 2013).

A Global/Nested version of the GFDL FV3 dynamical core (hfvGFS) was tested in near real-time and provided forecasts of tropical cyclone track, structure, and intensity during the active part of the 2017 Atlantic hurricane season. This high-resolution configuration of FV3 model forms the basis for HAFS developments discussed in section 14. The model domain covered the entire Atlantic basin with a horizontal resolution of 3 km (Fig. 24). As summarized in previous reports, the model's performance was comparable to the best operational intensity guidance (Hazelton et. al. 2018).

7 https://www.gfdl.noaa.gov/fv3/

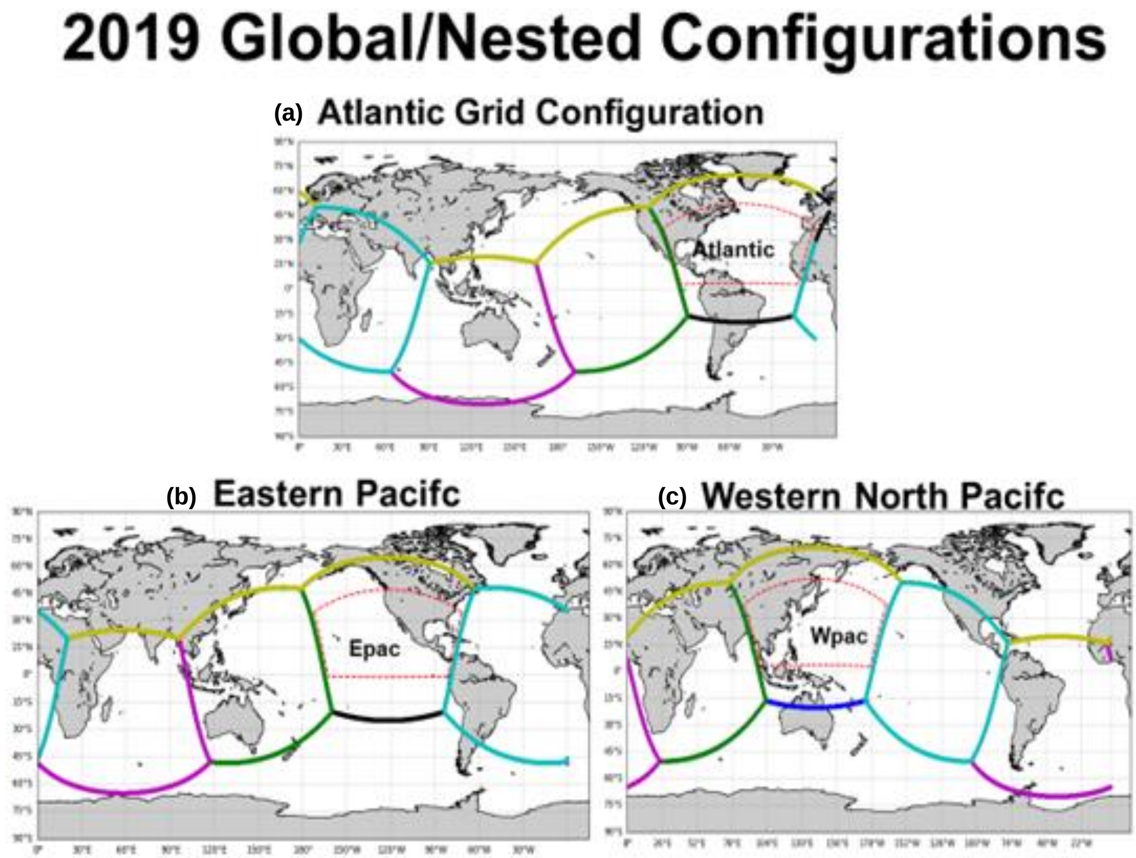

Figure 24: Proposed 2019 Global/Nested configuration for the Atlantic (top), Eastern Pacific (bottom left) and Western North Pacific (bottom right), highlighting the region to be covered by the 3km hfvGFS for each of the three domains.

In the 2018 version of hfvGFS, significant physics changes were made, which were developed at GFDL and incorporated into the experimental 13km global version of the model that is run on the Jet supercomputer in near-real time throughout the year. These upgrades included replacement of the GFS EDMF boundary layer scheme with the YSU scheme, introduction of a 1d ocean model, and changes to the model advection scheme. Although the intensity skill of the 2018 hfvGFS was reasonable and comparable to the HWRF through 3 days (Fig. 25), some of the physics upgrades introduced a significant high intensity bias which appeared most evident for sheared storms and degraded the overall season performance. After post-season analysis was done by GFDL and Hurricane Research Division (HRD) scientists, a modification to the micro-physics was added to hfvGFS and extensively tested (i.e., effect of the sedimentation heating) to address the high bias. As shown in Fig. 25, this physics change (purple line) led to reduced intensity errors at days 4 and 5, and dramatically reduced intensity bias compared to the 2018 version of the 3km model (red line). Testing and evaluation of this change in hfvGFS is presently continuing at both HRD and GFDL.

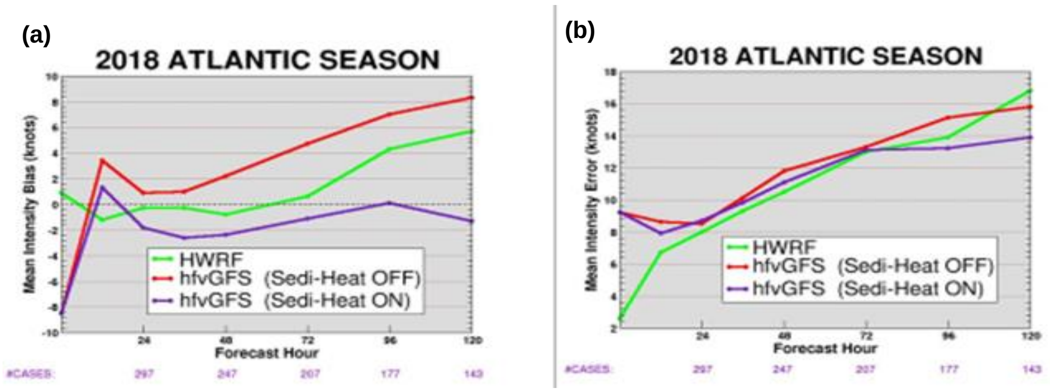

Figure 25: Homogenous comparison of intensity bias (left) and intensity error (right) for the 2018 Atlantic hurricane season using the 3km high-resolution hfvGFS for the 2018 near-real time configuration (red) and with modified micro-physics (purple), compared to the operational HWRF (green).

The hfvGFS model is also being evaluated in both the East and West Pacific (Fig. 26) with the 3km domain shifted west and centered in these 2 basins of interest. Plans are currently being formulated to run hfvGFS in these basins in near real-time for the upcoming 2019 season. For the limited number of cases tested so far (Fig. 26), the model performance is comparable with both HWRF (green) and COAMPS-TC (blue). Most noteworthy is the hfvGFS encouraging forecasts of RI, particularly for Hurricane Lane and Typhoon Mangkhut, where the hfvGFS did a significantly better job in correctly predicting the maximum intensity compared to HWRF and COAMPS-TC. Currently plans are being considered for the 3-km Global Nested configuration (as well as the stand-alone regional version) to be run in the Atlantic by the Environmental Modeling Center (EMC) in near real-time as the prototype versions of the HAFS, while HRD and GFDL run the 3km nest in the 2 Pacific basins, varying the basin of interest between the synoptic (0z, 12z) and off synoptic (6z, 18z) times or on a priority basis.

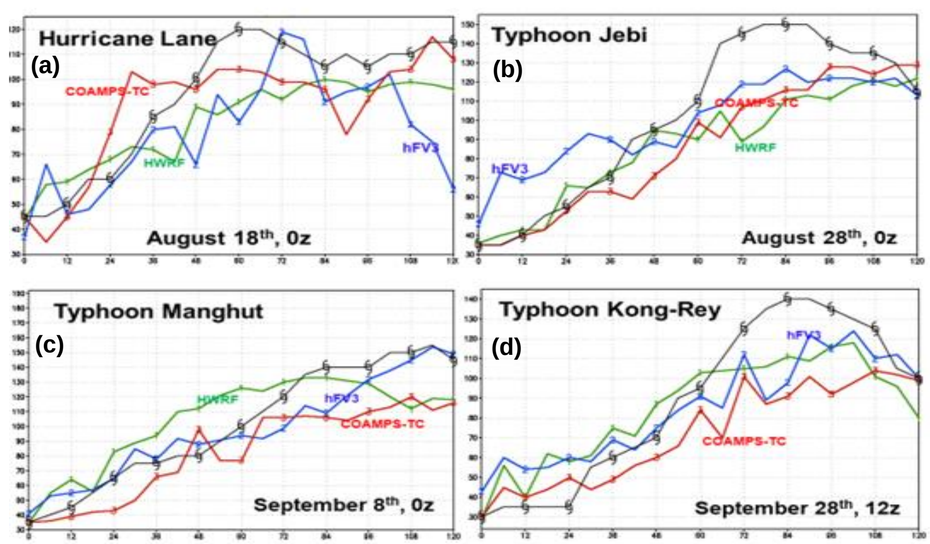

Figure 26: Intensity errors for four forecasts from the 2018 tropical seasons for Hurricane Lane (upper left) and three super Typhoons in the Western Pacific using the 3km hfvGFS (blue) compared to the operational HWRF (green) and COAMPS-TC (red).

Another important project in HFIP is the development of the moving nest for the HAFS system. Based on the success of the Basin-Scale HWRF - HRD, GFDL and EMC are working together to build the first ever global model with moving nest (section 13, Fig. 29).

9. New Products, Tools, and Services at NHC

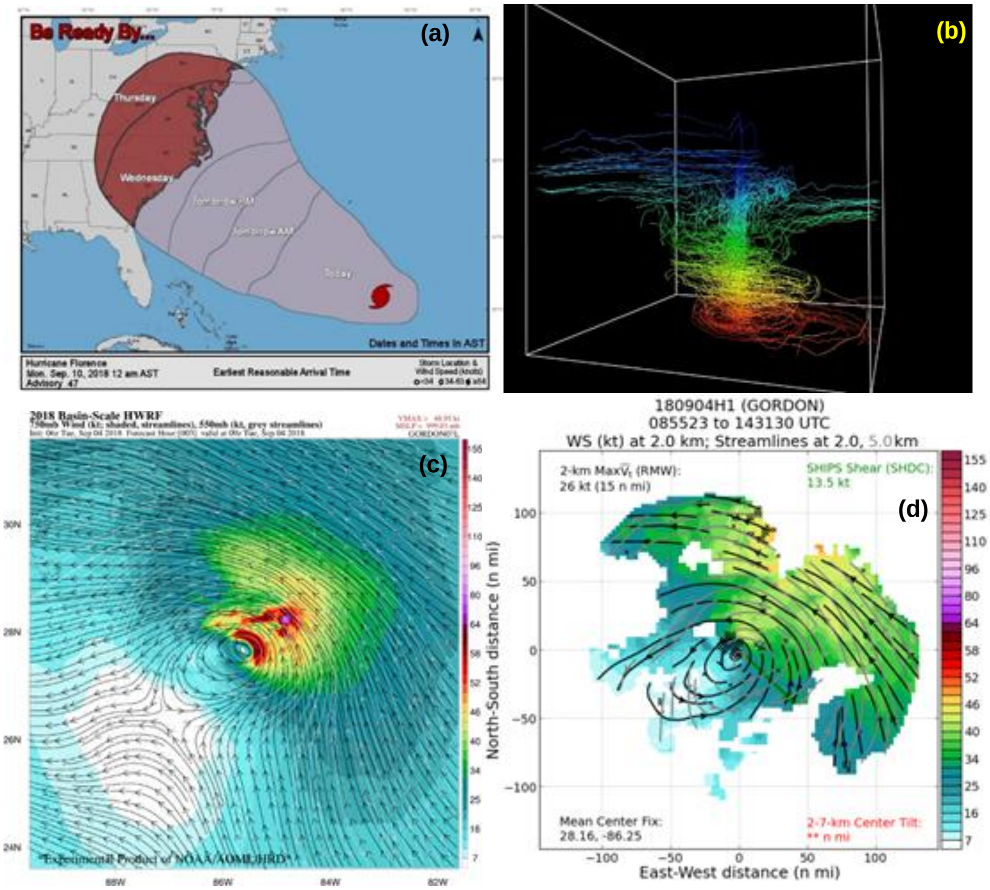

Figure 27: Examples of PPAV (Post Processing and Verification) products and results from 2018: (a) "Be Ready By" graphic used for media briefings during Hurricane Florence; (b) 3-D visualization of trajectories in a modeled HWRF storm; (c) Basin Scale HWRF analysis of Tropical Storm Gordon at 550 mb, colored contours are wind speed with streamlines plotted above; (d) Analysis of P-3 Tail Doppler Radar data for Gordon at 2 km using the same colored contours and streamline analysis as (c).

a. Improvements to NHC Public Products

The HFIP PPAV team effort in 2018 focused on support and enhancement of NHC public products, improvements to NHC's guidance suite, and model diagnostics. The Time of Arrival graphics were operationally implemented in 2018, and GIS files were made available to media and emergency management partners for the first time. This allowed emergency managers to access timing information directly in software that is used to make evacuation decisions. The "Be Ready By" graphic was automated and was used at NHC during media briefings prior to the landfall of Hurricane Florence (Fig. 27a). This graphic is a version of the Time of Arrival graphic customized for use in media briefings. In addition, improved post-storm visualizations of storm surge and other hazards were created for use in NHC reports and outreach efforts.

An upgrade to the wind speed probability was implemented in 2018 by improving the modification of the tropical cyclone wind field over land to better account for the decrease of winds of land areas. This resulted in a more realistic representation of the wind speed probabilities over land. In addition, research into computing ensemble tracks based on dynamical, rather than statistical, guidance continued to show improved skill over the current wind speed probabilities, and this will be considered for future operational implementation.

b. Improvements to Forecaster Guidance and Tools

The HCCA model has been a major achievement for the HFIP program. Further improvements to the model were made in 2018, including the migration of the code to the NWS operational supercomputing framework, the addition of Central Pacific storms for the Central Pacific Hurricane Center (CPHC) area of responsibility, real-time updates to the training dataset, and the evaluation of the HMON as an additional input. The HCCA model is frequently used by NHC forecasters, and was explicitly referenced in nearly 1/3 of NHC tropical cyclone discussions during the 2018 hurricane season.

Several improvements were made to other portions of the NHC guidance suite, including the statistical-dynamical models. SHIPS and LGEM began to incorporate GOES-16 data, their developmental databases were expanded through 2017, and wind radii was added to SHIPS and LGEM processing. Sensitivity and optimization studies for NHC's simple consensus aids were conducted to determine the models used operationally in 2018. Work on the Statistical Consensus model SPICE was also done to examine the replacement of GFDL with HMON. SPICE output will be used to investigate persistent biases in models such as the HWRF. In addition, the NHC targeting program was updated to help forecasters plan G-IV synoptic surveillance flights. Finally, several improvements were made to the user interface of the Automated Tropical Cyclone Forecast (ATCF) system. The upgrades allowed the Weather Prediction Center (WPC) to issue public advisories for inland depressions in the ATCF and NHC began to issue 48 h hurricane-force wind radii forecasts.

c. Improvements to Model Diagnostics and Visualization

The HFIP community also made progress in model diagnostics and visualization in 2018. At NHC, prototype 3-D visualizations were created to help forecasters evaluate model data (Fig. 27b). HRD produced visualizations of HWRF model storms broken down into cylindrical wavenumber components to help visualize the RI of hurricanes such as Michael. Additional work was done to improve diagnostics of storm vertical tilt and response to shear, and to create graphics that place model data and radar observations into a common analysis framework (Fig. 27c, 27d). Model graphics and diagnostics were again available on the HFIP website in real-time, and a few new models were added, such as the FV3-GFS.

Another public visualization tool, the NCAR NHC display was available in real-time for the 2018 season, and an internal version was installed at NHC and used for situational awareness. The NHC Display system key advancements developed during FY2018 include the development of the F-Deck and B-deck editing tool. Additional tools that have been added include a wind radii and a distance measuring tool. A new diagnostic feature has been added that color codes forecast based on forecast age. Several additional updates were implemented including a new TVCN membership, fix details on time series plots, and new plot features. The NHC Display system was used extensively by NHC during the 2018 Hurricane season and also for international training purposes during the 2018 WMO hurricane workshop. Several new model implementations were evaluated, including the FV3-GFS, the NWS 'National Blend of Models (NBM)', and potential upgrades to the HWRF. Evaluations of experimental models such as the Basin Scale HWRF (HWRF-B) were also undertaken to identify model biases. NHC assisted with the evaluation of data denial studies conducted by EMC to determine the impact of supplemental radiosondes and G-IV dropwindsondes on GFS track skill.

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 current 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 is easily transferable into operations. As of July 2018, the current code version of the HWRF system v4.0a is available for the HWRF community. Apart from US, there are about thousand HWRF model users in about 189 countries[8]. 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. User support was expanded with the Stream-2 efforts, the significant one being the Basin-Scale HWRF. This research system can support any number of high-resolution movable nests centered on TCs in either the Atlantic or eastern North Pacific basin. Working with HRD, the DTC also supported the transition of this research version to the latest community repository, enabling users to access all advancements in the HWRF system including the end-to-end Basin-Scale configuration (excluding ocean coupling and data assimilation). A similar testbed activity is recommended for transitioning the proposed HAFS.

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)	Evaluating Initial Condition Perturbation Methods in the HWRF Ensemble Prediction System

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

12. Socio-economic Aspects of HFIP

NOAA's National Hurricane Center'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.

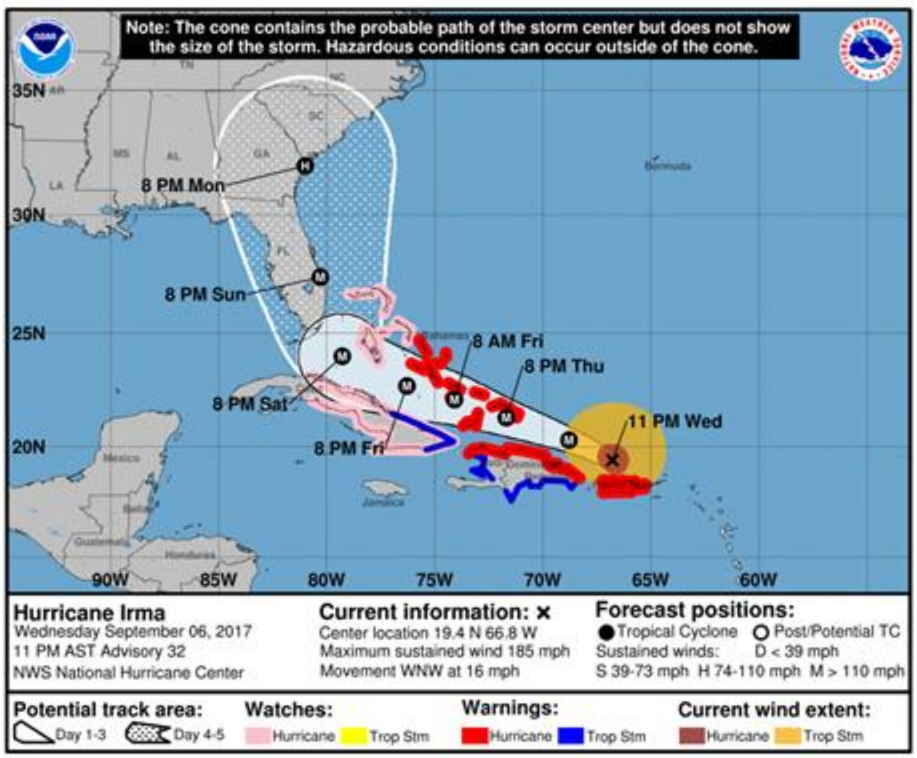

Figure 28: The 5-day cone of uncertainty with black track line toggled off.

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 public's, 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. HFIP State-of-the-art

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 increased confidence in those forecasts. Specific goals include reducing track and intensity errors by 20% in 5 years and 50% in 10 years and extending the useful range of hurricane forecasts to 7 days. HFIP has invested in improving hurricane forecast, high-performance computing (HPC) and in capacity building in hurricane research. Under HFIP, there have been significant improvements to NOAA's forecasts through improved modeling, observations and inner core data assimilation techniques resulting in increased accuracy in the numerical guidance for tropical cyclone intensity predictions.

Sustained HFIP investments in research and development (R&D) and HPC have led to the creation and transitions of the high-resolution HWRF system from research to operations (R2O). The system is now paving the way, around the globe, and removing the initial roadblocks associated with predicting intensity changes, dynamical prediction of which was nearly non-existent until 2009 (Fig. 2b). HWRF has improved by about 40-60% since 2011 over the Atlantic basin (Fig. 29).

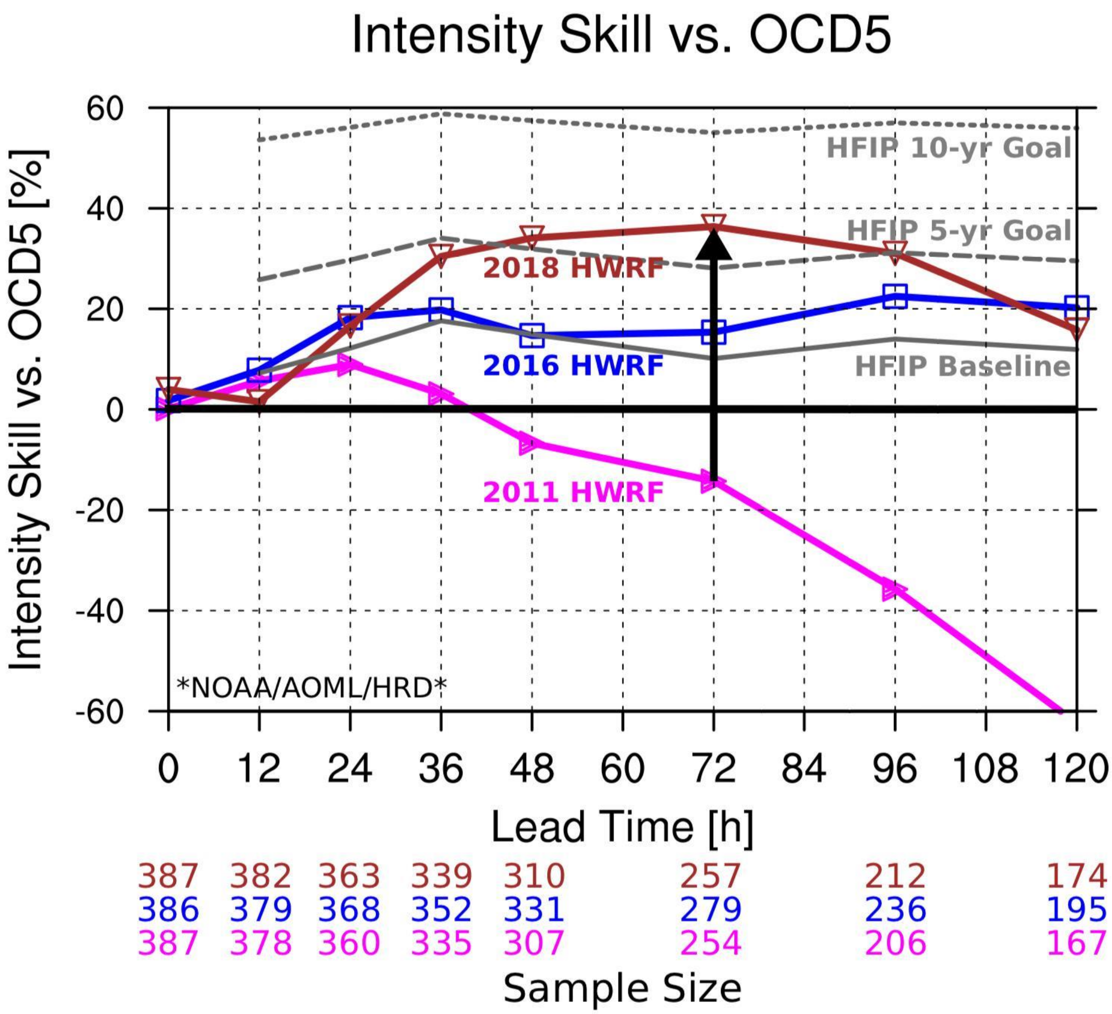

Figure 29: HWRF intensity skill relative to Decay-SHIFOR for 2011-2018 Atlantic seasons including 5-year and 10-year HFIP goals.

Since 2014, HWRF is run in operations in all global basins and is used by forecasters for reliable intensity guidance (Fig. 9 & Fig. 10) worldwide. Significant improvements to the HWRF system are attributed to a number of major changes since 2012, including a new, higher- resolution nest that is 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. This kind of development and T&E would not be possible without the support of the HFIP JET-HPC in Boulder that was dedicated for Hurricane R2O early in the program. HFIP has also build a capacity of model users, developers and hurricane scientists both within NOAA and academia to tackle the next generation hurricane forecast improvements. There were 5 Federally Funded Opportunities over the last 10 years for HFIP, awarding 40 grants to University PI, totaling $10.5M. All these HFIP efforts, has led to hundreds of publications related to HWRF within that period[9].

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

---

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

under HFIP (Fig. 22). 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. 22). Transitions of the multiple moving nested, HWRF to next generation global and regional prediction system within the unified forecasting system (FV3-UFS) is underway and expected to provide another step improvement 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, 2017 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 it now complements the HWRF efficiency to improve forecast consensus. 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. The HFIP JET-HPC has been critical for all HWRF/HMON advancements.

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 numerical guidance. Nevertheless, as shown in Figure 29, although HWRF has improved 40-60% in intensity predictions, that is still sufficient to only meet the 50% of targeted HFIP intensity goals (i.e., 20% reduction). Part of the reason may be associated with lack of progress with dynamical guidance until 2012. In fact, until 2012 intensity predictions lagged even the baseline (Figure 29) 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 Fig. 29 was carried out by calculating the median errors. The errors shown in Fig. 29 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.

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. For most part, the lack of improvement in the RI forecast skill, is rooted in our lack of understanding on 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) and cycle to cycle inconsistency (e.g., Hurricanes Harvey, 2018) makes the RI prediction still a very elusive problem. Some sustained HFIP research is recommended in this area. In the past 5 years, some other key challenges were reducing global modeling guidance errors to extend forecasts to 7 days with accuracy; improving utilization of high-resolution ensembles for model initialization and in product generation; better utilization of satellite observations in cloudy regions surrounding tropical cyclones; and identifying observations required to improve intensity forecasts. These areas are still under exploration in HFIP.

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 next generation of 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.

a. Hurricane Analysis and Forecast System (HAFS)

The major goals of the new weather act are planned to be addressed through the development of a multi-scale, multi-model system called Hurricane Analysis and Forecasting System (HAFS). The HAFS is NOAA's next-generation multi-scale numerical model and DA package, 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 within the framework of the UFS and its rolling three-year SIP. Central to the development of HAFS will be the FV3 dynamical core with embedded moving nest capable of tracking the inner core region of the hurricane at 1-2 km resolution.

One of the biggest successes achieved during the first phase of HFIP was the creation of the high-resolution HWRF system. HWRF, storm following, nested grid modeling system was designed to operate at a horizontal resolution of 2 km or less desired for capturing tropical cyclone inner core processes as well as the interactions with the large-scale processes, proven to be critical for improving track, intensity, rainfall and size predictions. It may be noted that HWRF has been providing only track and intensity guidance to forecasters until now. This system, with further advancements in ensembles, DA techniques and better use of hurricane observations, will be the starting point for the first version of HAFS. HAFS will evolve into 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 NGGPS/SIP objectives of the UFS.

The first HAFS developer's workshop was held on November 5, 2018 in Miami, FL. The primary objective of this meeting was to discuss the current state of operational hurricane capabilities and challenges, transition to the UFS-based FV3 system consistent with SIP with the development of HAFS, physics and DA advancements, set out NHC priorities, and build storm surge strategy. The results, accomplishments, and lessons learned from the 2018 hurricane season were discussed allowing the development of a multi-year strategy for improving hurricane forecast guidance. HFIP has begun investing in research advances to develop the HAFS by modernizing the global predictions system based on FV3GFS dynamical core while continuing research initiatives with the regional ensembles, telescoping two-way interactive nests, and probabilistic forecasts.

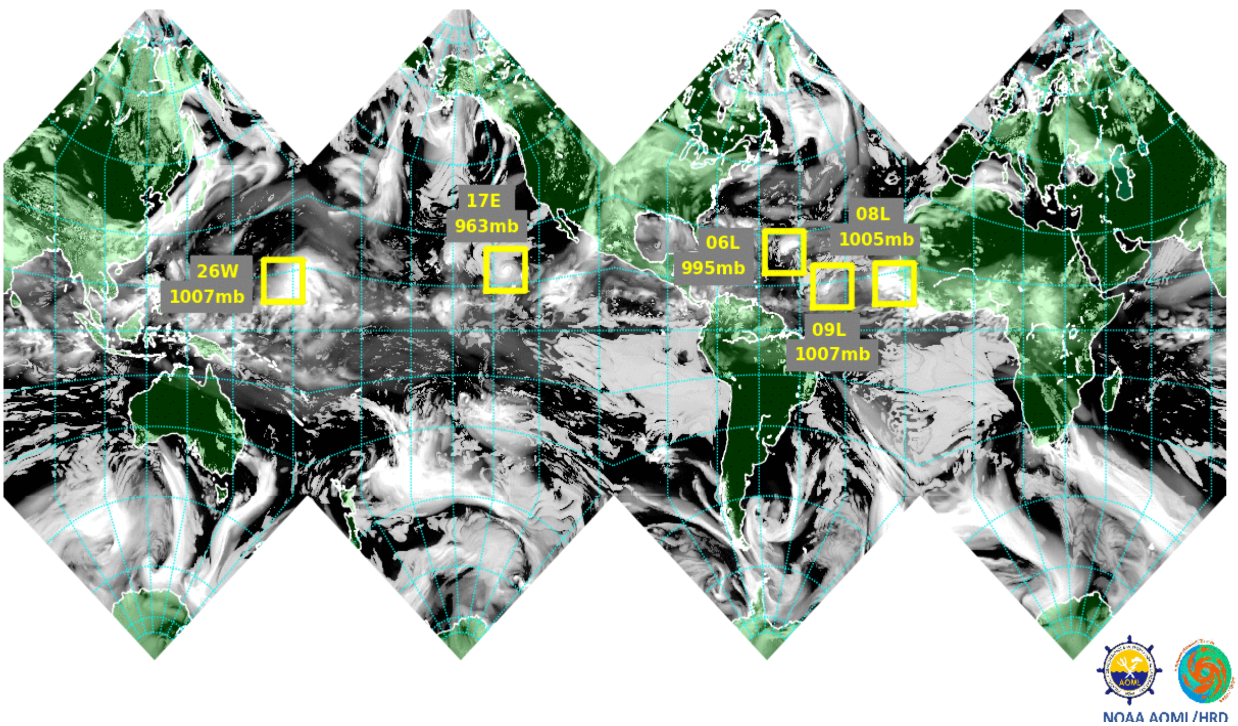

Figure 30: Cartoon from HFIP presentations showing how high-resolution nests may be moved seamlessly within the 6 faces of the FV3 cube sphere grid.

The FV3 model, itself is fully tested and cloud resolving, however, 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 and should be capable of following the tropical storms for several days. In addition, two-way interactive nests are essential for improving the accuracy of forecasts. AOML in joint partnership with GFDL and EMC is working on these developments to transition advances in HWRF to FV3-HAFS under hurricane supplemental.

b. Challenges

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.

Atlas, R., V. Tallapragada, and S. G. Gopalakrishnan, (2015): Advances in tropical cyclone intensity forecasts. Mar. Technol. J., 49, 149–160. https://doi.org/10.4031/MTSJ.49.6.2.

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., (2018): 2017 National Hurricane Center Forecast Verification Report, 2017 Hurricane Season. NOAA/NWS/National Hurricane Center, Annual Report, June 2018. [Available online https://www.nhc.noaa.gov/verification/pdfs/Verification_2017.pdf].

Cangialosi, J. P., (2018): National hurricane center 2018 forecast verification (preliminary). The 10th HFIP Annual Rev. Conf. presentation, Nov. 6, 2018, Miami, FL. [Available online at http://www.hfip.org/events/annual_meeting_nov_2018/docs/2_0900AM_Cangialosi.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. 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., (2018): 2017 HFIP R&D Activities Summary: Recent Results and Operational Implementation, HFIP Technical Report: HFIP 2018-1. [Available online at http://www.hfip.org/documents/HFIP_AnnualReport_FY2017.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]

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
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
ECMWF	European Centre for Medium-range Weather Forecasts model
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
GDAS	Global Data Assimilation System
GEFS	Global Ensemble Forecast System
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 Project
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
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
NCEP	National Centers for Environmental Prediction
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
NOGAPS	Navy Operational Global Atmospheric Prediction System
OAR	Oceanic and Atmospheric Research
OFCL	Official National Hurricane Center Forecast
POD	Probability of Detection
RI	Rapid Intensification
RW	Rapid weakening
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
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