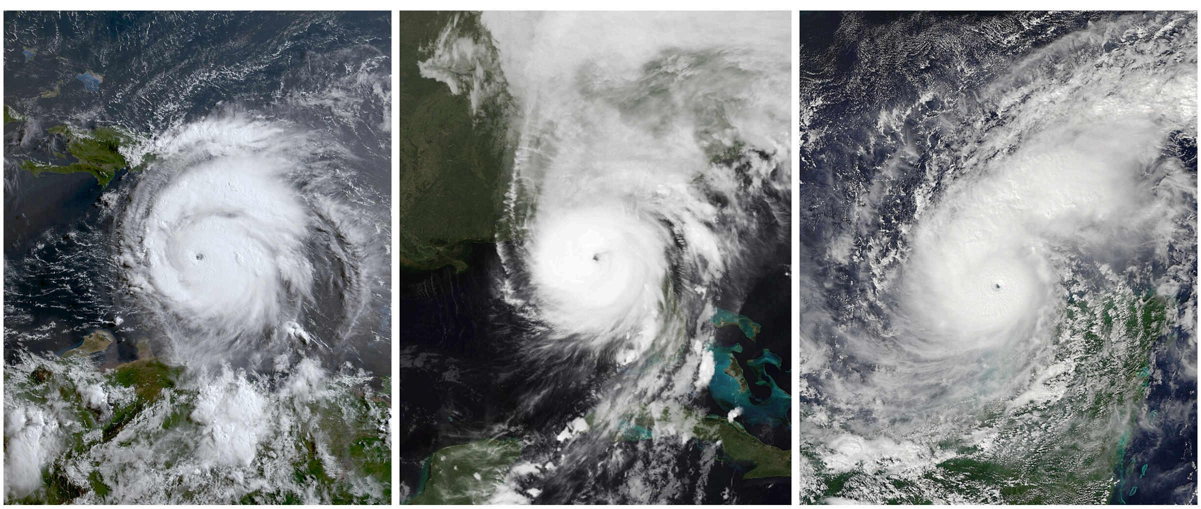
NOAA

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
United States Department of Commerce

NOAA HURRICANE FORECAST IMPROVEMENT PROJECT

Hurricane Forecast Improvement Program (HFIP) Annual Report 2024

NOAA technical memorandum HFIP-2024
https://doi.org/10.25923/yc7m-vw53
4 August 2025



Cover: GOES-16 geocolor image of (left) Hurricane Beryl at 1230 UTC 2 July 2024, (center) Hurricane Helene at 0146 UTC 27 September 2024, and (right) Hurricane Milton at 1920 UTC 07 October 2025. Images courtesy of NOAA/NESDIS/STAR.

Annual Report Writing Team

Aaron Poyer
William Komaromi
Jason Anderson
Sundararaman Gopalakrishnan

Ghassan Alaka
Michael Brennan
Jordan Dale
Andrew Hazelton
Wallace Hogsett
Bin Liu
Shirley Murillo
Kathryn Newman
Jiayi Peng
Jamie Rhome
Pablo Santos
Vijay Tallapragada
Xuejin Zhang
Zhan Zhang

Table of Contents

Executive Summary

This technical report describes the activities and results of the Hurricane Forecast Improvement Program (HFIP) that occurred in the 2024 hurricane season. This year's report provides key highlights of progress and achievement that the program has made in 2024 as well as the future direction and development.

In 2024, the HFIP program actively engaged in significant briefings and outreach initiatives with NOAA's management, executive leadership and Congressional committee, reinforcing the program's commitment to transparent communication and alignment with organizational priorities. The major focus of this report is to highlight the success of upgrading the Hurricane

Analysis and Forecast System (HAFS) to version 2 on July 16, 2024. Additionally, this report will examine the success and challenges that HAFSv2 encountered with certain storms during the 2024 season. We will also provide an overview of the annual meeting, results of various real time experiments, and highlight publications that came out in 2024 related to HFIP work and development.

The 2024 North Atlantic hurricane season was above average, with record breaking activity occurring after an unusual peak season lull. There were 18 named storms, of which 11 developed into hurricanes, with 5 of those becoming major (category 3+) hurricanes. The season was extremely active for the U.S. coastline, with 5 hurricane landfalls (Beryl, Debby, Francine, Helene, and Milton). The end result was a particularly deadly and damaging hurricane season, with 399 documented fatalities and $129.7 billion in damages, making the 2024 hurricane season the third costliest on record. Hurricane Helene was particularly significant due to its extended damage swath from straight-line wind damage and freshwater flooding hundreds of miles inland, across the interior Southeastern U.S. and the southern Appalachian Mountains. In contrast, the 2024 eastern North Pacific hurricane season was quieter than normal, with 14 named storms, of which 5 developed into hurricanes and 3 became major hurricanes. Across the NHC area of responsibility, 12 tropical cyclones underwent rapid intensification (RI), defined as an intensification of 30 kt or more in 24 hours, from 9 tropical cyclones in the Atlantic basin (Beryl, Debby, Francine, Helene, Isaac, Kirk, Milton, Oscar, Rafael), and 3 tropical cyclones in the Eastern North Pacific (Gilma, John, Kristy).

The major highlights of 2024 were:

1. HAFSv2 launched on July 16, 2024, with improved vortex initialization (VI) and data assimilation (DA), increased horizontal resolution from 2 km to 1.8 km, several physics upgrades, and integration of the more advanced MOM6 ocean model to replace HYCOM. Overall, HAFSv2.0 features an improved intensity forecast at days 4-5, improved track forecast at days 1, 4, and 5, and improved storm structure and wind radii prediction at both short and long lead times.

2. A strategic plan writing team was assembled, and a draft Strategic Plan 2025-2035 with revised 5-year and 10-year goals and objectives was produced. The new strategic plan was discussed at the HFIP Annual Meeting 2024, and revisions were incorporated. At the time of this report, the updated strategic plan is under management review.

3. The real-time experimental HAFS Ensemble was run on the Azure cloud in 2024. The 2024 HAFS Ensemble was a notable improvement from 2023, with an increase in the number of ensemble members from 21 to 31, an increase in horizontal resolution from 6 km to 4 km, an increase in the number of vertical levels from 65 to 81, as well as various updates to physics and ocean coupling. The 2024 HAFS Ensemble produced overall superior track and intensity performance, improved RI prediction, and a better spread-skill relationship versus the 2023 version of the ensemble.

4. HFIP Real-time Experiments (HREx) 2024 featured a number of scientific and technological advancements related to HAFS and/or complementary systems, including the basin-scale HAFS multistorm (HAFS-M) configuration and GSL's HFIP-MPAS for hurricanes. Many of these experiments showed promise and improved performance versus HAFS v2.0, and components of some of these experiments will be integrated into the HAFS v2.1 upgrade in 2025.
5. There is a continued need to address the day 1-3 HAFS intensity skill gap versus the legacy HWRF and HMON systems, which are scheduled for retirement prior to the 2026 hurricane season.
6. Looking forward to 2025, based upon feedback and discussions with the HFIP Executive Oversight Board (HEOB) and at the HFIP Annual Meeting 2024, HFIP seeks to produce revised executive governance and a roadmap of the end-to-end "value chain". These actions will help to maximize HFIP's value and impact to stakeholders and end users, including forecasters, emergency management, and the general public.

1. History of HFIP

1.1. Introduction

This report describes the Hurricane Forecast Improvement Program (HFIP), its goals, proposed methods for achieving those goals, and the most recent results from the program, with an emphasis on advances in the skill of operational hurricane forecast guidance. Section 1 of this report describes the background, goals, and baselines for measuring success within the HFIP program. Section 2 focuses upon the initial operating capability (IOC) of the Hurricane Analysis and Forecasting System (HAFS), highlights high-resolution hurricane modeling successes from the 2024 hurricane season, and highlights experimental and developmental versions of the model, including a HAFS ensemble, that are in the testing and evaluation stages for possible future transitions. Section 3 highlights the engagement of HFIP with the community and summarizes the HFIP Annual Meeting 2024. Section 4 summarizes this report, and previews a new direction for the future of HFIP that will be elaborated upon in further detail in the upcoming HFIP Strategic Plan 2025-2035. For more background information, readers are referred to earlier reports available on the HFIP website.

1.2. The Hurricane Forecast Improvement Program (HFIP)

Originally established as the Hurricane Forecast Improvement Project, authorized in 2007 and beginning in 2009, HFIP was created within NOAA in response to the particularly damaging landfalling hurricanes (e.g., Charley, 2004; Wilma, Katrina, Rita, 2005) in the first half of that decade. HFIP's original 5-year (for 2014) and 10-year goals (for 2019) were to:

- Reduce average track errors by 20% in 5 years, and by 50% in 10 years for days 1-5
- Reduce average intensity errors by 20% in 5 years, and 50% in 10 years for days 1-5

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

For more than a decade, HFIP has been providing the unified organizational infrastructure, funding, and compute resources for NOAA, university, and private partnerships to coordinate the hurricane research needed to achieve the above goals, improve storm surge forecasts, and accelerate the transition of model codes, techniques, and products from research to operations. HFIP focuses on multi-organizational activities to research, develop, demonstrate, and implement enhanced operational modeling capabilities, dramatically improving the numerical forecast guidance made available to the NHC, as well as enhancing the interpretation of that guidance. The success of HFIP over its 15-year history can be easily visualized in terms of the reduction in 48-hour forecast track error, a lead time at which hurricane watches have typically already been issued and the NHC is preparing to begin issuing warnings. The average consensus track forecast error has been reduced from 110 mi in 2007, to 75 mi in 2017, to 55 mi in 2023 (Fig. 1). Over-warning areas and unnecessary evacuations have also been correspondingly reduced, which saves would-be evacuees time and money, and also reduces evacuation traffic congestion. Through HFIP, NOAA continues to improve the accuracy of hurricane forecasts, with applied research using advanced computer models.

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

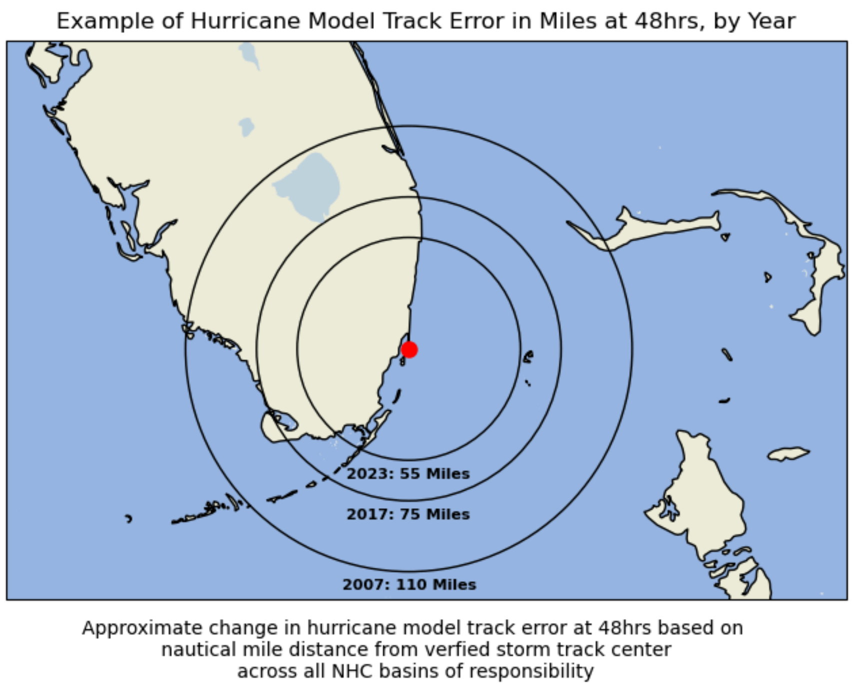


Figure 1: Average model forecast consensus track errors at 48 h lead time in 2007 (110 mi), 2017 (75 mi), and 2023 (55 mi), centered over south Florida for reference.

In 2017, Congress passed the Weather Research and Forecasting Innovation Act including Section 104, reauthorizing HFIP as the Hurricane Forecast Improvement *Program*. Under HFIP, this Congressional Act instructed NOAA to maintain a project to improve hurricane forecasting with the goal of developing and extending accurate hurricane forecasts and warnings in order to reduce loss of life, injury, and damage to the economy. HFIP has a particular focus on improving the prediction of rapid intensification and track of hurricanes, improving the forecast and communication of surges from hurricanes, and incorporating risk communication research to create more effective watch and warning products. In response to this charge, the HFIP Strategic Plan 2019-2024 was updated outlining the research and development needed to continue improving hurricane forecast guidance, enhance probabilistic hazard products, and design a more effective tropical cyclone (TC) product suite to better communicate risk to the public and emergency management community. Under the updated plan, HFIP will continue to address the original goals of reducing track and intensity forecast errors by 20% within 5 years and 50% within 10 years, and to extend forecasts out to 7 days, particularly with focus on rapid intensification guidance. In addition, the updated plan extends HFIP's purview to improving guidance on predicting storm structure and all hurricane hazards (surge, rain, associated severe weather, gusts as well as sustained winds) at actionable lead times for emergency managers (e.g., 72 hours). Improved hazard guidance will derive from dynamical model ensembles enabling probabilistic hazard products and improved track, intensity change and structure (radii to maximum and 35-knot winds) predictions before formation and throughout the storm's life cycle. Using social science research, HFIP will design a more effective tropical cyclone product suite to better communicate risk and transition all current tropical hazards products.

One of the key strategies defined in the revised hurricane forecast improvement strategic plan in response to the proposed framework for addressing the Weather Act of 2017, is to advance an operational HAFS. HAFS is a multi-scale model and data assimilation package capable of providing high-resolution analyses and forecasts of the inner core structure of the TC out to a lead time of 7 days, which is key to improving size and intensity predictions, as well as the large-scale environment that is known to steer TCs and provides favorable/unfavorable dynamic (e.g., vertical wind shear) and thermodynamic (e.g., mid-tropospheric moisture) conditions. HAFS will provide an operational analysis and forecast system out to 7 days for hurricane forecasters with reliable, robust and skillful guidance on TC track and intensity (including RI), storm size, genesis, storm surge, rainfall and tornadoes associated with TCs. It will provide an advanced analysis and forecast system for cutting-edge research on modeling, physics, data assimilation, and coupling to earth system components for high-resolution TC predictions within the UFS. HAFS development has been supported under several supplemental appropriations, or “supplementals”, including the (i) 2018 Improving Forecasting and Assimilation (IFAA) Portfolio, (ii) 2019 Improving Forecasting of Hurricanes, Floods, and Wildfires (IFHFW), (iii) 2022 Disaster Relief Supplemental Act (DRSA), and (iv) 2022 Bipartisan Infrastructure Law (BIL), Provision 3.

HFIP model development is organized along both an operational and an experimental line of activities. Operational support includes the final code approval, transition of code from the Environmental Modeling Center (EMC) to NCEP Central Operations (NCO), and the use of the operational models by NHC hurricane specialists issuing their forecasts. For operationalization, only the final version of models and products are running and code is optimized to minimize the computational cost. Experimental model development, performed under HFIP Real-time Forecasting Experiments (HREx), includes the testing and evaluation of a variety of different variations of the code, in both real-time demonstrations and retrospective experiments. For this work, code is not yet fully optimized since it is still in flux, and multiple experimental variations of a similar model may be running in parallel to determine which version demonstrates superior performance. The purpose of HREx is to demonstrate that the application of advanced and innovative science, technology, and increased computing will lead to the desired increase in accuracy, and other improvements in forecast performance. Due to the significant computational cost associated with performing demonstration experiments, which can be 3-10x larger than operational computing requirements, HFIP leverages a number of internal research and development high performance computing (HPC) systems that are managed by NOAA, including: Jet at the David Skaggs Research Center (DSRC) in Boulder, CO, Hera at NOAA Environmental Security Computing Center (NESCC) in Fairmont, WV, and Gaea at Oak Ridge National Laboratory (ORNL) in Oak Ridge, TN. HFIP also supports testing and development work performed at the external Mississippi State University (MSU) HPC systems, Orion and Hercules.

2. HFIP in 2024

This section summarizes the activities and results of HFIP that occurred in 2024. The major focus of this section is the testing, operational implementation, and real-time performance of the Hurricane Analysis and Forecast System (HAFS) version 2, along with the continued development of the experimental HAFS Ensemble run on a cloud computing environment. We begin this section with some background on the deadly and costly 2024 Atlantic and East Pacific hurricane seasons, and also summarize some recent advances in social, behavioral, and economic sciences (SBES) related to HFIP.

2.1 Background on the 2024 Hurricane Season

The 2024 Atlantic hurricane season was highly impactful for the United States, as well as internationally across the Caribbean, Greater and Lesser Antilles. The season featured five U.S. landfalls, including: Beryl (Cat 1), Debby (Cat 1, and second landfall as a tropical storm), Francine (Cat 2), Helene (Cat 4), and Milton (Cat 3). The season also featured five international landfalls, including: Beryl (Cat 4 and Cat 2), Ernesto (Cat 1), Oscar (Cat 1), and Rafael (Cat 3). There were 436 documented fatalities, making the season the deadliest since 2017, and approximately $130 billion in preliminary damages, making it the third costliest season on record.

In the East Pacific, Hurricane John was a powerful category 3 that made landfall on the west coast of Mexico. John stalled off the coast of Mexico and re-intensified, making a second landfall and producing tremendous rainfall totals of up to 37" in Guerrero, and similar totals in Oaxaca and Michoacan. John resulted in 29 fatalities and approximately $2.45 billion in damages.

Preliminary fatality reports from the 2024 Atlantic hurricane season are consistent with recent decadal trends: fatalities due to storm surge continue to fall, while fatalities due to freshwater flooding have remained relatively steady (Fig. 2). Since the early years of HFIP (and even before), significant funding and efforts have been devoted to improved modeling of storm surge, in addition to warning and communication efforts. While there have been focused efforts to improve flood and flash flood predictive lead times, continued high fatalities from tropical cyclone induced freshwater flooding suggests a need for further research, development and evaluation is needed to close gaps in predictive skill and communicating threats.

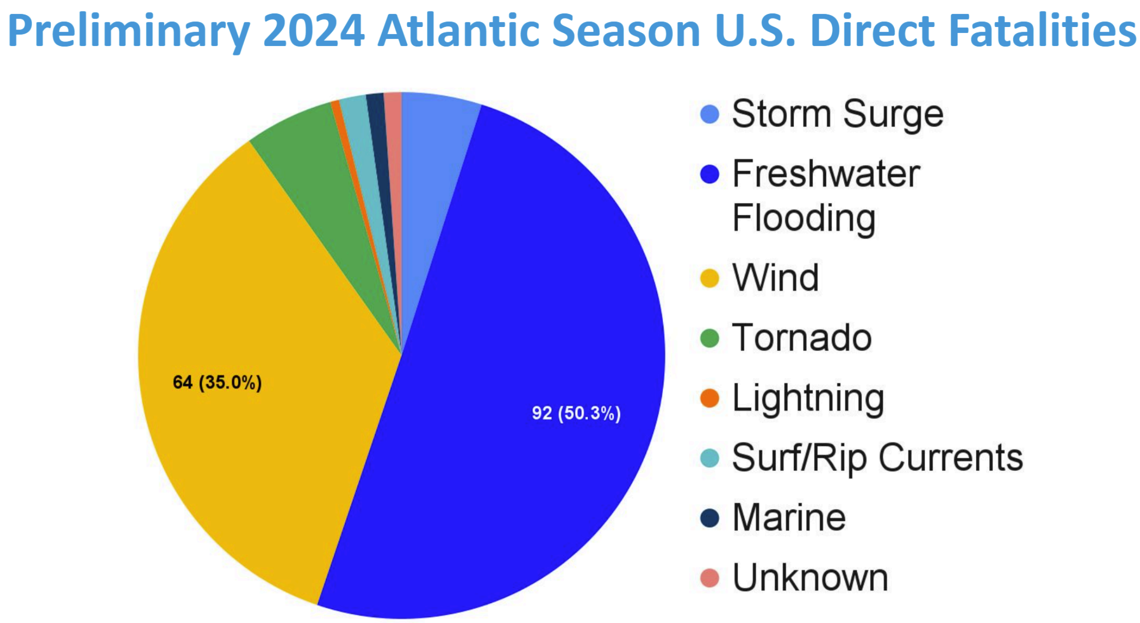
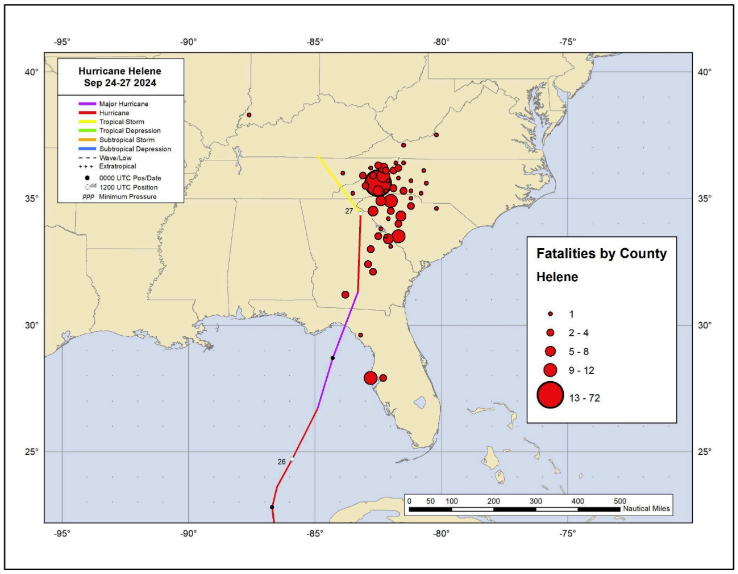


Figure 2: preliminary causes of U.S. direct fatalities and causes in the 2024 Atlantic hurricane season.

Of all the tropical cyclone impacts in 2024, Hurricane Helene was both the deadliest and the most damaging in the U.S., with 224+ direct fatalities and causing approximately $75 billion in damages. Whereas most TC fatalities typically occur in the vicinity of landfall, before the storm begins to weaken, Helene was somewhat unique amongst recent TCs in that the fatality distribution was notably skewed inland along the southern end of the Appalachian Mountains of NC and SC, as well as GA, well inland of the FL panhandle landfall (Fig. 3).

2024 Atlantic Season - Helene



- 224 Fatalities (Preliminary)
- Fatalities by State (direct+indirect)
  - NC: 119
  - SC: 50
  - GA: 23
  - TN: 17
  - FL: 12
  - VA: 2
  - IN: 1

Figure 3: Geographic distribution of preliminary direct fatalities from Hurricane Helene (2024).

Hurricane Milton, while not quite as deadly or destructive as Helene, was also noteworthy in its own right due to an associated tornado outbreak that occurred over the FL peninsula, spawning 45 tornadoes including 3 of EF3 intensity. This outbreak of tornadoes became the largest single day of tornadoes in FL state history, surpassing Hurricane Irma (2017), resulting in 6 fatalities and greater than $500 million in damages. Similar to the flash flood risk from Helene, tornadoes spawned by Milton demonstrate the continued importance of “all hazards prediction” for TC risk communication and preparedness.

Polling indicates that the American public is willing to pay more than $500M annually for planned hurricane forecast improvements (Molina et al., 2021). It has also been shown that dollars saved from forecast improvements (e.g. reduced area of coastline evacuated ahead of storms) since the inception of HFIP far exceed dollars spent. The reality remains that dollars spent on response and recovery continue to far outweigh dollars spent on preparedness and mitigation by a factor of 50:1 to 100:1 (Fig. 4).

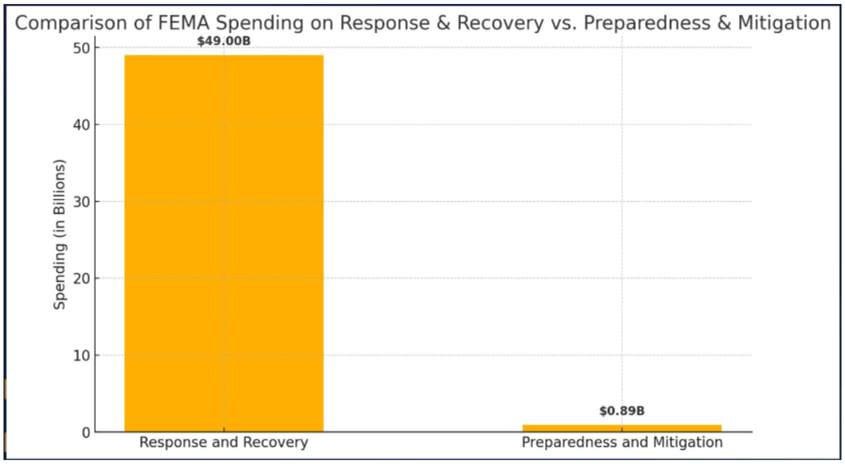


Figure 4: Comparison of FEMA spending on response and recovery versus preparedness and mitigation from preliminary preparedness and response spending from November 2024 (courtesy Dulce Suarez, FIU).

2.2 NHC Forecast Verification

The 2024 hurricane season presented a number of forecast challenges, but was overall an impressive season for NHC forecast performance. It was a phenomenal year for track prediction, with record low forecast errors set at all lead times (Table 1). There was a sharp decrease in track error relative to the 2023 spike, and the long-term decrease in track error remains pronounced (Fig. 5a). However, the degree of difficulty in predicting TC track varies from season to season. As such, NHC also normalizes their track forecast skill with respect to a climatological model, CLIPER5. Even when normalizing with respect to climatology, NHC's skill scores in 2024 remain near the top of the chart, albeit not number one at all lead times (not shown). However, it is worth noting that there was a steady rise in track skill with respect to CLIPER5 from 1990 up until 2015, but track skill has been relatively stable since then. This suggests that we are beginning to approach the limits of predictability with traditional NWP models. Alternatively, we may need to explore novel techniques to more effectively utilize ensembles and AIWP to restart the upward trend.

The long-term trend in intensity error in NHC's official forecast is quite noisy, albeit with a gradual downward trend (Fig. 5b). The year-to-year variation in intensity forecast error is greater than for track error, perhaps due in part to the smaller scale of processes involved, and due to the fact that a single missed RI or RW event can contribute to significant errors for a season. Overall, intensity errors in 2024 were slightly above 2023, but below long-term averages (Table 1).

Forecast Hour	Sample Size	Track Error (n mi)	Intensity Error (kt)

12	270	20.3	5.1
24	241	29.2	7.7
36	214	38.0	9.8
48	190	47.3	11.3
60	167	56.8	11.8
72	143	66.9	12.6
96	105	89.0	13.6
120	75	115.0	13.2

Table 1: NHC mean track forecast error (n mi) and intensity error (kt) at various lead times from 12 h to 120 h in 2024 in the Atlantic basin.

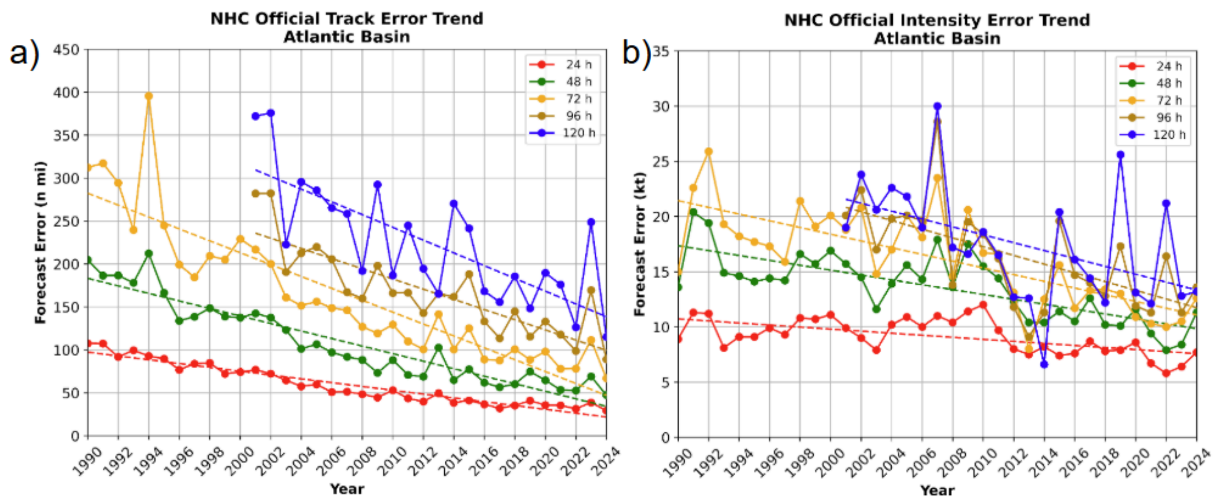


Figure 5: NHC official (a) track and (b) intensity forecast error at 24, 48, 72, 96, and 120 h lead times in the Atlantic basin from 1990-2024.

Amongst sources of forecast guidance, including regional and global NWP and blended consensus aids, NHC performed as well or better than all individual guidance for track, while blended consensus aids such as HCCA and TVCA came in a close second place (Fig. 6a). The top individual model from 12-72 h was the ECMWF deterministic (EMXI), while the GFS (GFSI) performed the best from 84-120 h. All hurricane regional models are in the middle of the pack.

For intensity, the NHC official forecast is also as good or better than any individual model or consensus aids, followed by consensus aids HCCA, NNIC, and IVCN competing for second place with the exact leader varying by lead time (Fig. 6b). Amongst individual NWP models, HMON (HMNI) took the top spot for intensity, once again proving the difficulty of the new HAFS variants in out-performing the legacy systems. HWRF, HAFS-A, and HAFS-B all performed somewhat similarly, trailing HMON. Statistical models also demonstrated their continued worth after all of these years, with DSHP and LGEM outperforming dynamical models from 96-120 h. While global models continue to become higher and higher resolution over time, and are somewhat more realistic for TC intensity and structure than they used to be in years past, the GFS (GFSI) and ECMWF (EMXI) were not competitive with regional or statistical models for intensity in 2024. Additionally, it should be noted that years of investment by HFIP (and other sources) in improving RI prediction have paid off, as NHC's performance in predicting RI in 2024 was amongst the highest it has been, with a success ratio of ~0.60 and a probability of detection of ~0.65 (Fig. 7).

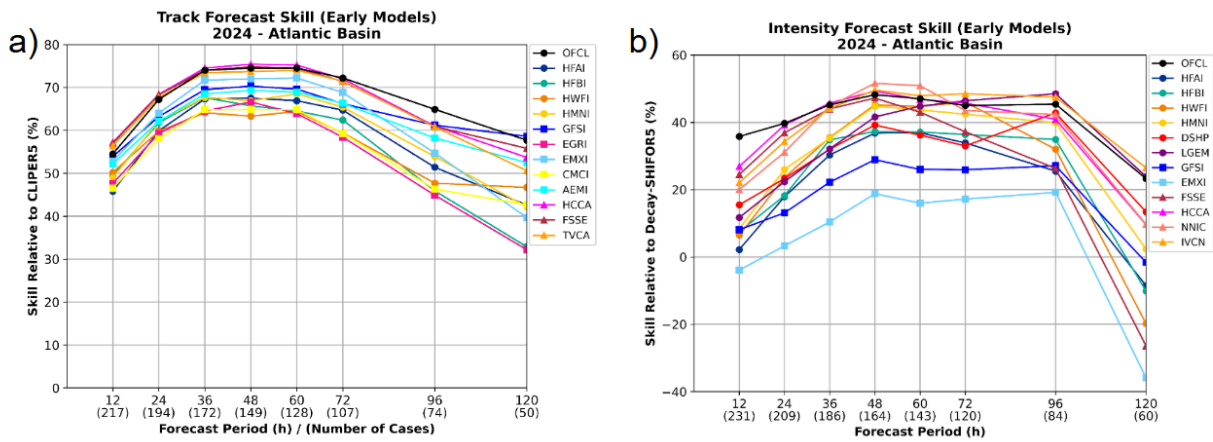


Figure 6: (a) Track and (b) intensity forecast skill as a function of lead time for the NHC official forecast (OFCL) and a variety of global models, regional models, and consensus aids in 2024 for the Atlantic basin.

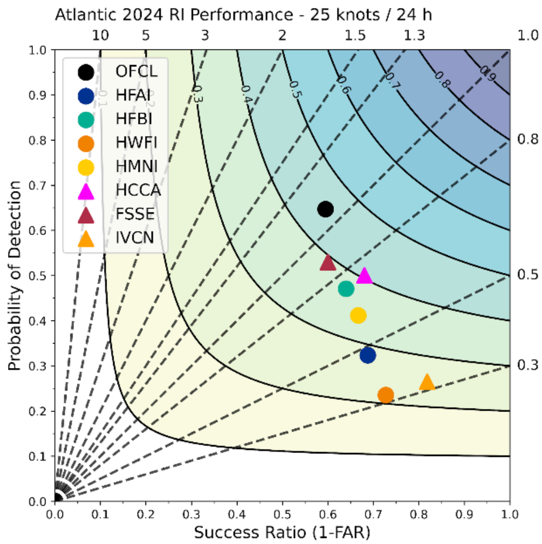


Figure 7: RI forecast verification in 2024 for the Atlantic basin, including the NHC official forecast (OFCL), regional models, and consensus aids.

The Eastern Pacific basin was quieter than average in 2024. Accordingly, there were fewer forecasts issued by NHC than usual. For track, the FSSE blended consensus model was the best model overall, and outperformed the NHC official forecast at most lead times (Fig. 8a). Quite similar to the results in the Atlantic, the best individual model for track at shorter lead times (36-48 h), while the GFS (GFSI) was the top model from 60-120 h. HAFS performed better for track in the Eastern Pacific in 2024, with HAFS-B (HFBI) the top performer for track amongst regional models. The UKMET (EGRI) and CMC (CMCI) models did not perform well in this basin this season.

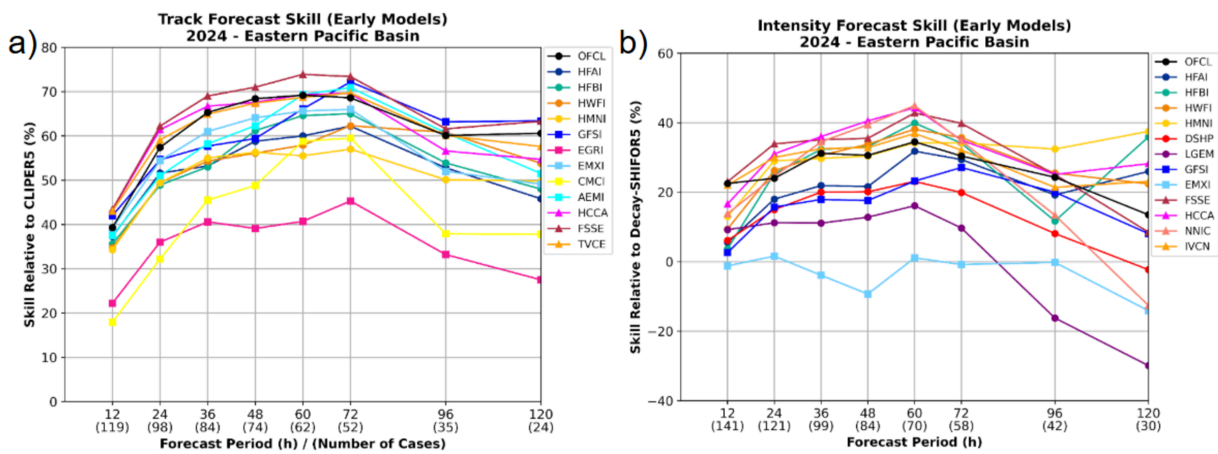


Figure 8: (a) Track and (b) intensity forecast skill as a function of lead time for the NHC official forecast (OFCL) and a variety of global models, regional models, and consensus aids in 2024 for the Eastern Pacific basin.

For TC intensity in the Eastern Pacific, blended consensus aids also outperformed NHC (Fig. 8b). Amongst individual models, including statistical models, the regional dynamical models performed the best. Overall HMON (HMNI) took the top spot for intensity in the Eastern Pacific this season, as it did in the Atlantic as well, again demonstrating the continued utility of the legacy systems. Interestingly, the GFS (GFSI) performed competitively with skillful statistical models such as DSHP for intensity, while LGEM and the ECMWF (EMXI) lagged behind.

2.3 HAFSv2.0

The first version of the Hurricane Analysis and Forecast System (HAFSv1.0) became operational at NCEP Central Operations (NCO) in 2023 under the UFS modeling framework and featuring the FV3 core, replacing HWRF as NOAA's flagship hurricane model, a major milestone for HFIP. One of the major accomplishments of HFIP in 2024 was the development, testing, and operational transition on 16 July 2024 of HAFSv2.0. The 2024 upgrade was notable in that it featured an increase in horizontal resolution from 2 km to 1.8 km, improved vortex initialization (VI) and data assimilation (DA), several physics upgrades, and integration of the more advanced MOM6 ocean model to replace HYCOM. Overall, HAFSv2.0 features an improved intensity forecast at days 4-5, improved track forecast at days 1, 4, and 5, and improved storm structure and wind radii prediction at both short and long lead times. One significant feature of the HAFSv2.0 upgrade was an improved VI scheme, which produced significantly improved realism of the TC structure (Fig. 9)

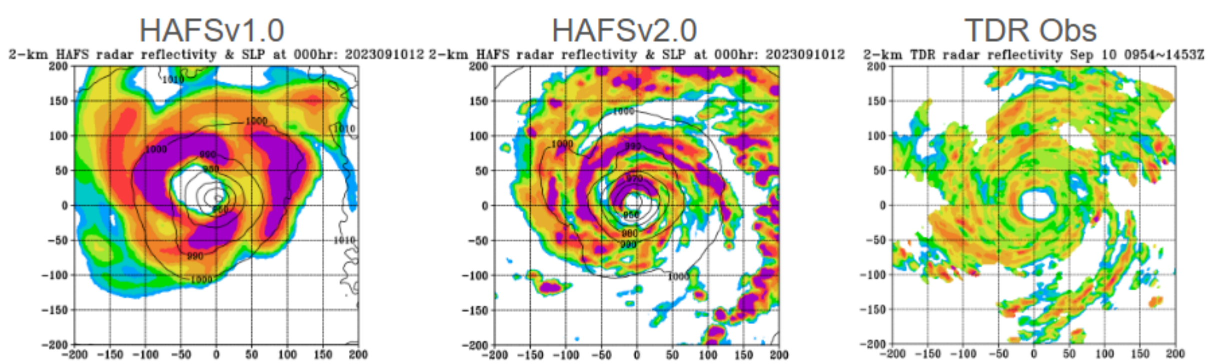


Figure 9: Simulated reflectivity at 2 km (dBz, shaded) and MSLP (hPa, contoured) from HAFSv1.0 (left), and HAFSv2.0 (center), compared with P-3 Tail Doppler Radar (TDR) observations from Hurricane Lee on 10 September 2023 at 12 UTC.

In the 2024 hurricane season, HAFSv2.0 performed competitively with other top global and mesoscale model guidance, albeit with some challenges. For TC track in the Atlantic, the GFS model had the highest skill, followed by HMON, and then HAFS-B, HAFS-A, and HWRF which all scored similarly (Fig. 10a). Track skill scores in the East Pacific were more favorable for HAFS, with GFS, HAFS-A and HAFS-B all competing for top skill overall at various lead times

(Fig. 10b). For intensity in the Atlantic and the East Pacific, the legacy HWRF and HMON systems outperformed HAFS-A and HAFS-B, particularly in the day 1-3 lead times (Fig. 10c,d). Day 1-3 intensity prediction challenges for HAFS were noted by NHC during the HFIP Annual Meeting in 2024, and is one of the primary reasons for NHC's advocacy towards not retiring the legacy models. Overall HAFS performance in 2024 was slightly lower than in the 3-year retrospective period used as verification prior to operational transition. However, the nature of the general circulation pattern across the basin, the synoptic environment near any given storm, and the somewhat chaotic nature of convection, which all vary considerably from year-to-year and from storm-to-storm, it is not surprising that there are noteworthy fluctuations in model performance.

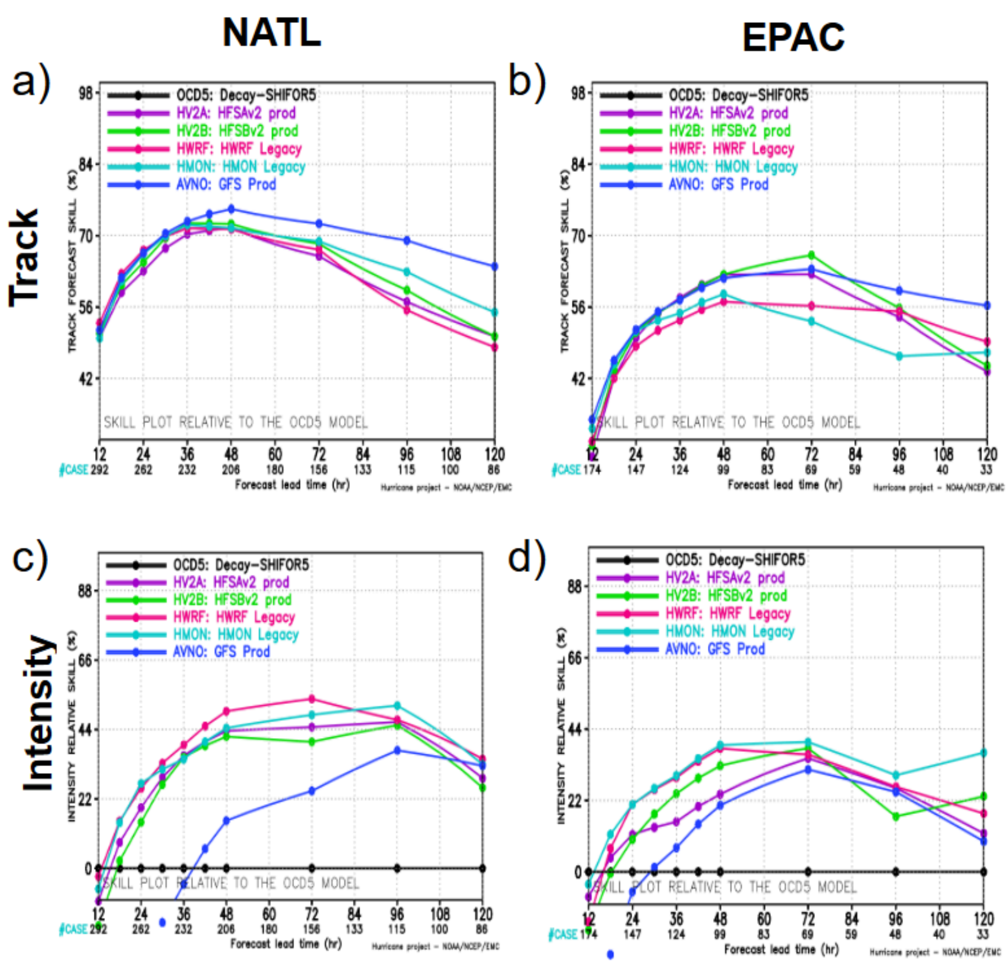


Figure 10: Performance of HAFSv2.0 -A and -B models versus the legacy HWRF and HMON systems, the GFS, and the OCD5 statistical baseline for skill for (a,b) track, and (c,d) intensity, in the (a,c) Atlantic and (b,d) Eastern Pacific basins.

HFIP is also particularly interested in low-frequency, high-impact scenarios such as rapid intensification. As such, RI in 2024 was also verified as a subset of intensity forecast skill. Since a model may or may not predict RI, and RI may or may not occur in reality, we verify RI using verification metrics such as probability of detection (POD) and success ratio (SR). For RI events in the north Atlantic, HAFS-B and HWRF were the top performers, with HAFS-A and HMON overall less skillful (Fig. 11). Amongst the models, HAFS-B had the highest POD, while HWRF had the highest SR.

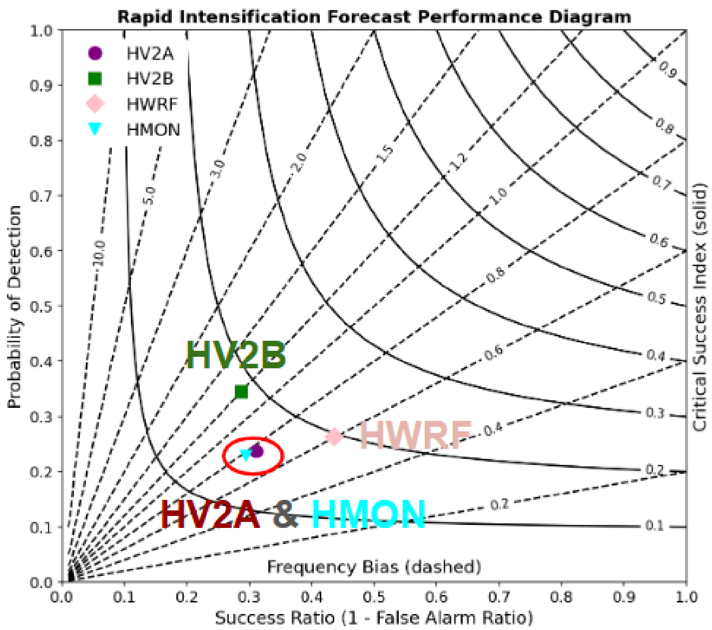


Figure 11: Rapid intensification performance diagram, probability of detection versus success ratio (1 - False Alarm Ratio), for HAFS v2.0 -A, -B, HWRF, and HMON.

2.4 HFIP Real-time Experiments (HREx) 2024

As has been the case since the Jet supercomputer came online in 2009 to be the workhorse for HFIP experiments, the HFIP Real-time Experiments (HREx) serves as a testbed for new and emerging modeling techniques through a series of real-time experiments. At the end of each hurricane season, discussions led by EMC evaluate and determine which experiments will be transitioned to operations for the following hurricane season. The decision to transition new modeling components is ultimately determined by three primary factors:

- Performance - whether or not the new capabilities improve upon forecast skill scores, reduce errors, or improve (reduce) biases, versus existing NWP or forecast techniques
- Readiness level - whether or not the new technology and/or coding is mature enough to run in operations
- Computational cost - whether or not the current operational configuration (core count, memory, disk space, etc) can support the new technology

In 2024, there were 4 HREx experiments run in real-time on Jet: HAFSv2.0.1A, HAFSv2.0.1B, HAFSv2.0.1M (Multistorm), and HFIP-MPAS. A short description of each experiment follows.

a) HREx 1: HAFSv2.0.1A

Led by EMC, the HAFSv2.0.1A is a variation of the currently operational HAFSv2A. Key model upgrades in this experiment include: improved vortex initialization (VI) using an adjusted minimum central pressure algorithm, newly developed storm-following Three-Dimensional Incremental Analysis Update (3DIAU) in inner-core DA, updates to the scale-aware SAS convection and to the PBL parameterization related to the background diffusivity, use of the Exponential-Random cloud overlap approach in Rapid Radiative Transfer Model (RRTMG), and air-sea interaction and coupling upgrades, including an upgrade to parallel Real-Time Ocean Forecast System (RTOFS) v2.5 input data and a switch to using Modular Ocean Model 6 (MOM6) ePBL mixed layer scheme.

Collectively, these changes resulted in unambiguously positive improvements to forecast skill in a three-year (2022-2024) retrospective sample. Track skill in the Atlantic basin was improved at all lead times in v2.0.1A by 3-7% (Fig. 12a), while intensity forecast skill was a net positive at most lead times, and as much as an 8% improvement at 96 h (Fig. 12b). Rapid intensification prediction was also improved in v2.0.1A versus v2A in terms of both probability of detection and success ratio (Fig. 12c). Given these highly beneficial improvements to HAFS-A with minimal effect on compute cost, HAFSv2.0.1A will serve as a baseline for the 2025 HAFS-A transition.

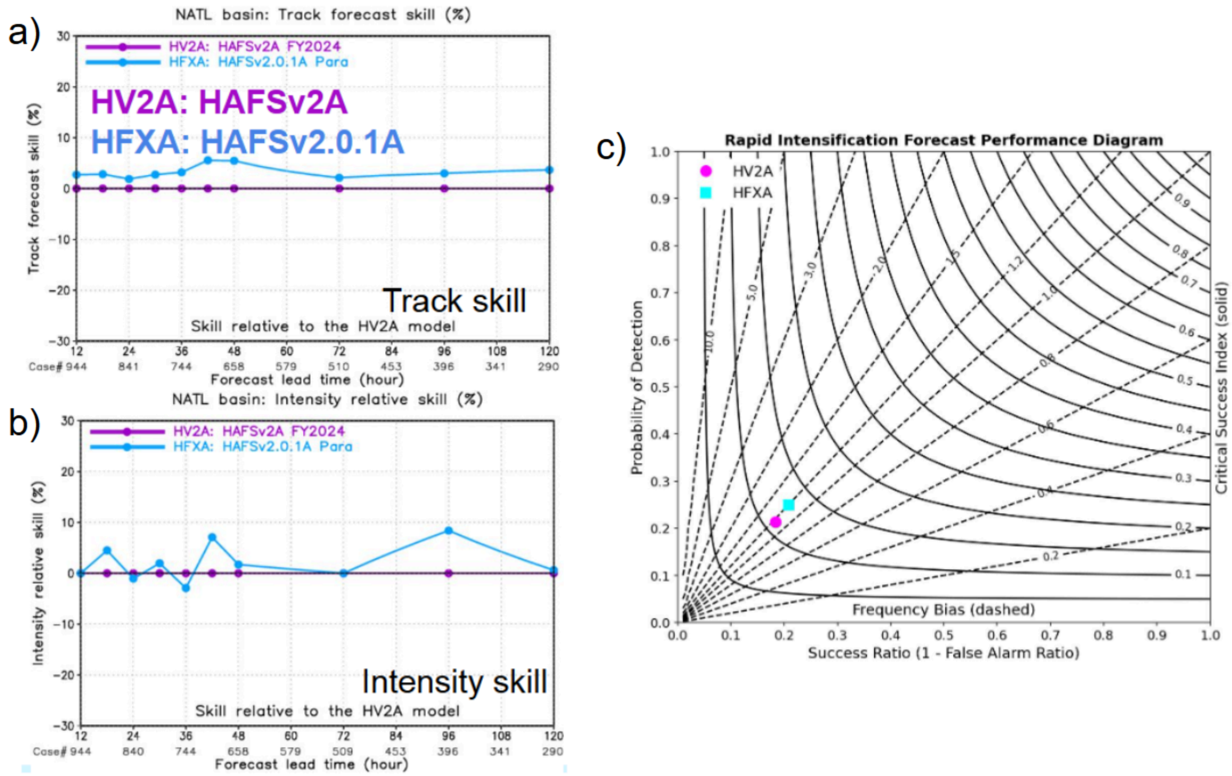


Figure 12: Forecast skill for (a) track, (b) intensity, and (c) RI performance comparison between HAFSv2.0.1A (magenta) and HAFSv2A (blue) in the Atlantic basin over a three-year (2022-2024) retrospective period.

b) HREx 2: HAFSv2.0.1B

The HAFSv2.0.1B experiment was a variation of the operational HAFSv2B, co-led by a combination of EMC and AOML/HRD. This experiment featured a variety of upgrades, including: changes to the scale-awareness (to increase parameterized updrafts, reduce downdrafts) and changes to detrainment and entrainment coefficients in the Tiedtke convection, modifications to the mixing length in the EDMF-TKE to offset the negative intensity bias associated with Tiedtke, updates to MOM6 ocean coupling with different mixing options, as well as several updates to initialization, including improved VI to reduce the TC central pressure bias, cloud relocation in DA cycling, and 3DIAU (also featured in the HAFSv2.0.1A experiment).

Overall, the HAFSv2.0.1B experiment showed promising performance improvements relative to operational HAFSv2B across 13 Atlantic TCs run in 2024. Improved day 5 track skill and day 4-5 intensity skill were consistent across the sample of cases. The experimental model performed well for structure across various forecasts for a number of significant TCs, including Hurricane Milton. Comparing 72-h forecast model simulated reflectivity and wind at 2-km (Fig. 13a,b) to analyses from the P-3 tail Doppler radar (Fig.

13c,d) from 09 October 2024 for Milton, it is apparent that the model was correctly capturing the hurricane's tiny pinhole eye as well as a subsequent eyewall replacement cycle (ERC), although in this particular forecast the ERC onset was slightly earlier than observed so the secondary eyewall is a bit more mature in HAFS than it was in reality. While the 3DIAU and VI improvements lead to an improved model initialization, unfortunately they do not appear to improve the problematic (as identified by NHC) day 1-2 intensity forecast yet; perhaps additional work and tuning is required. Overall these results demonstrate a strong starting point for the 2025 HAFS-B model upgrade, but may require running some additional test cases and tuning during the hurricane "off season".

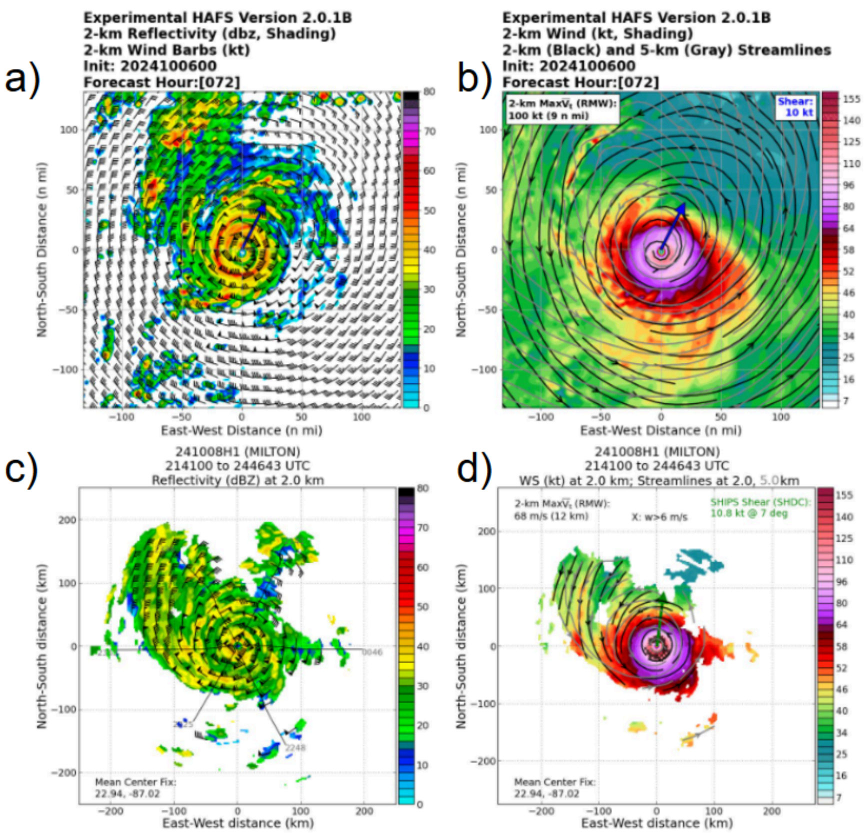


Figure 13: Hurricane Milton HAFSv2.0.1B 72-h forecast (a) 2-km altitude model simulated reflectivity (dbz) and wind barbs (kt), and (b) 2-km wind speed (kt), 2-km and 5-km streamlines, and deep-layer vertical wind shear vector, from the 00 UTC 06 October 2024 forecast valid 00 UTC 09 October 2024. P-3 tail Doppler radar (TDR) analyses of (c) 2-km reflectivity (dbz) and wind barbs (kt), and 2-km wind speed, 2- and 5-km streamlines, and deep-layer vertical wind shear vector, valid 00 UTC 09 October 2024.

c) HREx 3: HAFSv2.0.1M (Multistorm)

The HAFSv2.0.1M or “HAFS-M” experiment led by HRD constitutes a significant change in the HAFS grid and nesting configuration. The basin scale configuration encompasses the vast majority of the TC activity in NHC’s area of responsibility (AOR), including the entire Atlantic and much of the Eastern Pacific basins (Fig. 14). Unlike the 2024 operational HAFS, where each storm is run independently, the multistorm capability allows for multiple moving nests communicating with the same parent grid and with each other. This allows for more realistic storm-storm interactions, from binary interaction for TCs that are in close proximity, to remote cross-basin interactions via Rossby wave packets disturbing the waveguide. As an added bonus, the multistorm configuration is also more computationally efficient when there are 3 or more storms in NHC’s AOR.

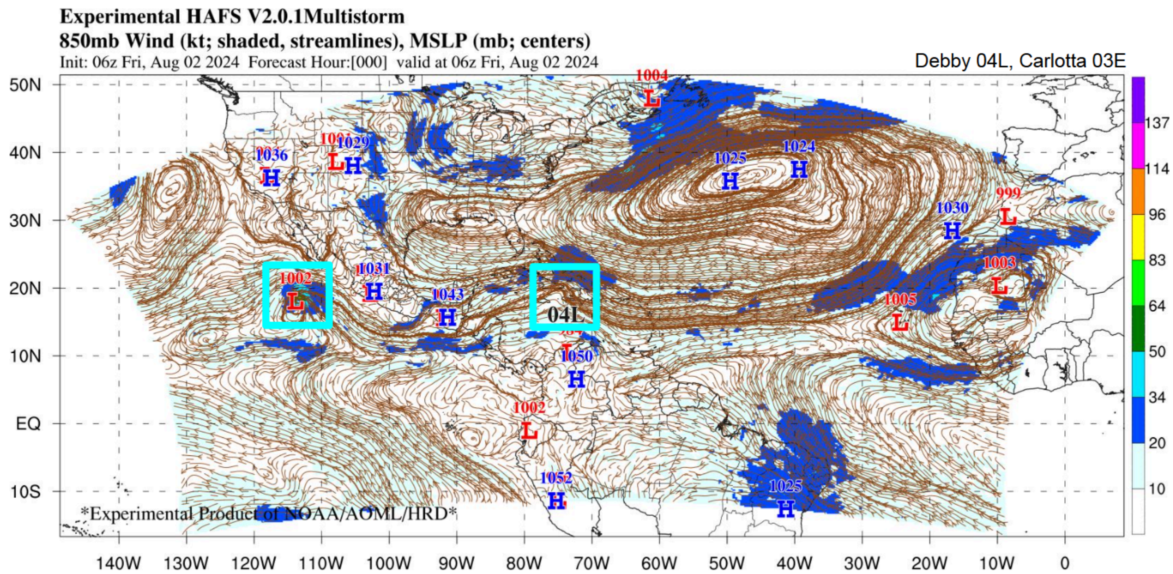


Figure 14: The HAFSv2.0.1M parent domain with two storm-following nests: one in the Atlantic and one in the East Pacific, valid 06 UTC 02 August 2024.

In addition to improved realism and computational efficiency advantages, the HAFS-M experiment also resulted in significant improvements in forecast skill for track (Fig. 15a), intensity (Fig. 15b), minimum central pressure, and wind radii at almost all lead times. Track improvements of 10-20%, and intensity improvements of 15-25%, constitute truly significant improvements which often require multiple rounds of model upgrades across several years. The possibility of seeing these returns from a single model implementation has the potential to be a significant achievement within the 15-year history of HFIP. The initial degradation in track and intensity at t=0 seen in the verification skill scores was likely due to a bug in the initialization that has since been addressed. Currently the only limiting factor preventing HAFS-M code from transitioning in 2025 is its readiness level, as the HAFS workflow is not *quite* ready to run operationally.

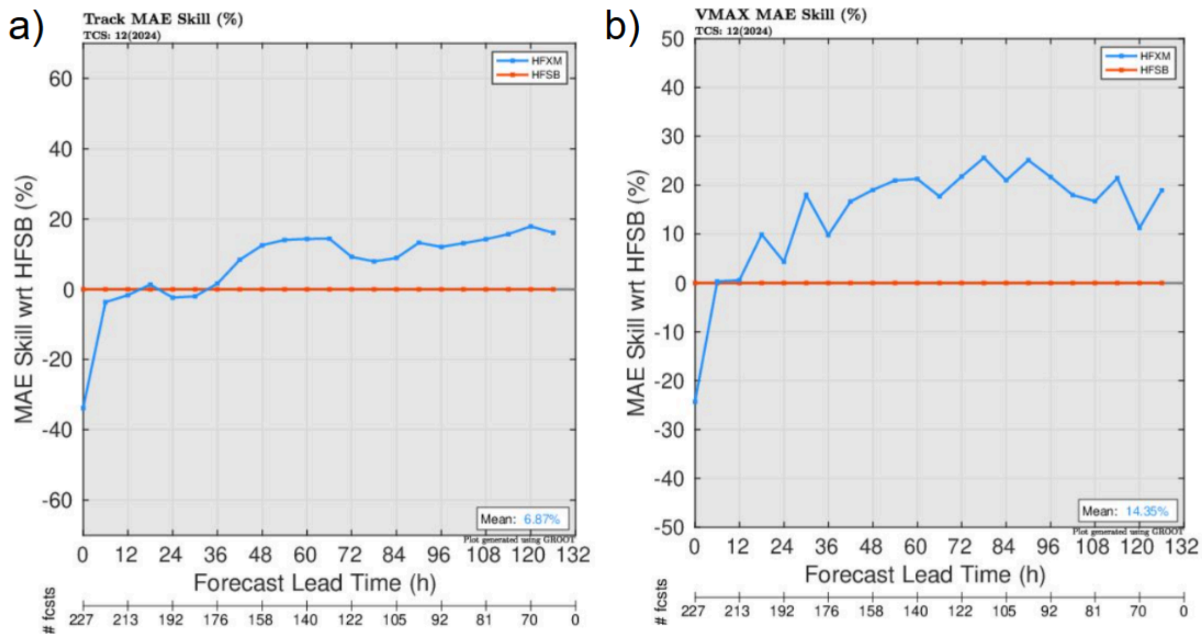


Figure 15: HAFS-M percent skill improvement versus HAFS-B for (a) track and (b) intensity from 0-120 h for 12 TCs in 2024.

d) HREx 4: HFIP-MPAS

The fourth and final HREx experiment in 2024 was performed by GSL. Unlike the other three HREx experiments which ran off the UFS FV3 core, GSL's HFIP-MPAS experiment ran off of NCAR's MPAS core. GSL and NCAR have demonstrated superior realism and predictive skill of MPAS versus FV3 for certain mid-latitude continental convection scenarios. This experiment is partially motivated by the desire to explore whether or not similar advantages extend to TC prediction. As a proof-of-concept, the HFIP-MPAS experiment for TCs ran at 3-km horizontal resolution with 80 vertical levels, using GSL's physics suite from RRFSv1 (with a few modifications). Since the workflow is new and relatively immature, no ocean coupling was used, and forecasts were cold-started (no DA or cycling).

Despite the simplified workflow, performance of the HFIP-MPAS experiment was quite impressive. Track forecast skill is comparable to or better than HAFS-A, HAFS-B, and GFS from 12-96 h, before degrading at 120 h (Fig. 16a). Intensity forecasts struggle at 0-24 h lead times without DA, but otherwise have comparable or better skill versus the NHC official forecast and NOAA's statistical and dynamical modeling suite from 36-120 h (Fig. 16b). It should also be noted that storm structure was also well predicted by HFIP-MPAS. As an example, Milton's eyewall replacement cycle on 09 October 2024 was well forecast by HFIP-MPAS 48-60 h in advance.

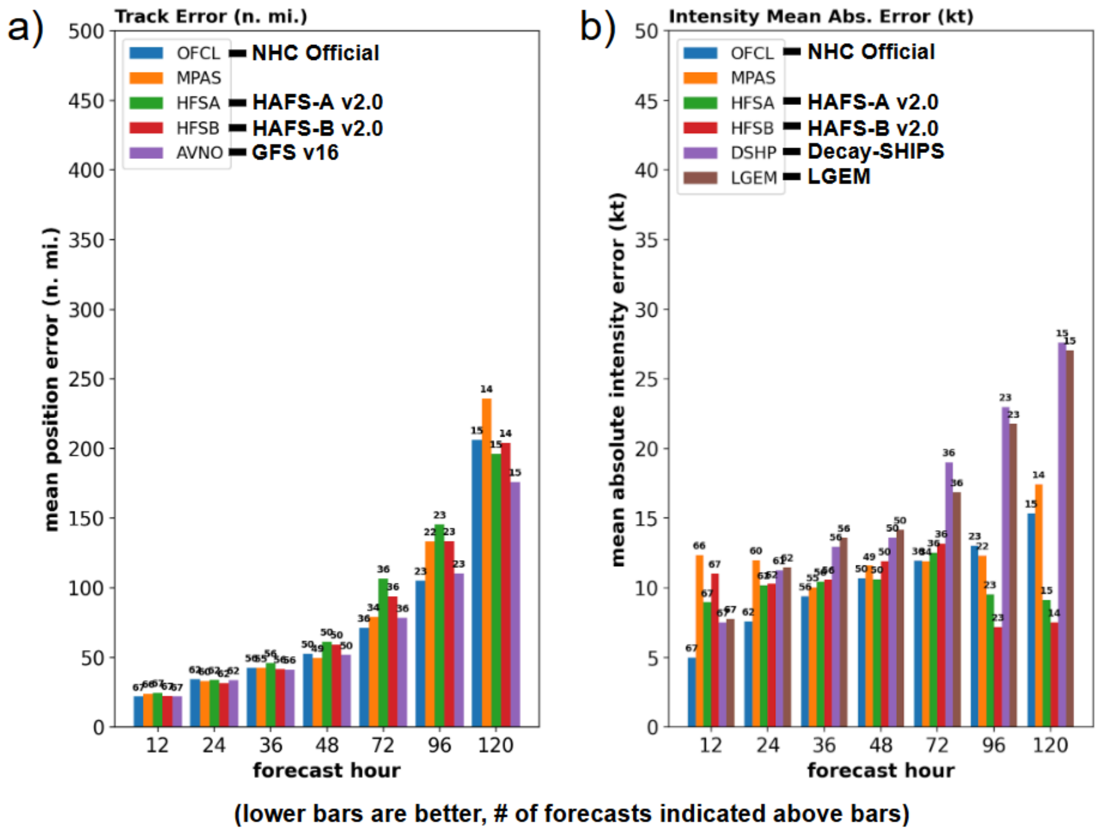


Figure 16: HFIP-MPAS forecast verification from 2024 for (a) track, and (b) intensity, in the Atlantic basin, constituting 67 total forecasts across 18 TC cases.

GSL anticipates being able to support real-time cycled data assimilation with MPAS and JEDI in 2025, and are considering a larger parent domain for HREx 2025. While this experiment is currently at too low a readiness level for transition in 2025, it has been an impressive proof-of-concept that warrants continued development over the next several years.

2.5 HAFS Ensemble

One of the significant strategic thrusts within the NWS has been to embrace probabilistic forecasts, and to implement improved graphics and products which convey probabilistic messaging, as outlined in the NWS Strategic Plan 2023-2033 and remains a “Ken’s 10” initiative. The experimental HAFS Ensemble run on the Azure cloud in 2024, also dubbed “HERC” (HAFS Ensemble in Real-time on the Cloud), provided probabilistic guidance to forecasters, which was under experimental review for evaluation in near real-time at NHC’s Hurricane and Ocean Testbed (HOT). The HAFS Ensemble in 2024 was a significant upgrade from 2023 (Table 2), with the number of members increased from 21 to 31, horizontal resolution increased from 6km to 4km, number of vertical levels increased from 65 to 81, microphysics was upgraded from the single-moment five-category GFDL scheme to the more sophisticated double-moment Thompson scheme, and half of the ensemble members were run using MOM6 ocean coupling instead of HYCOM for increased spread in ocean surface conditions. The ensemble continues to run with perturbed initial and boundary conditions to provide forecast spread, in addition to SPPT, SKEB, and SHUM perturbation schemes, which run during model integration. Additionally, a stochastic parameter perturbation (SPP) was added to perturb the ocean surface roughness in 2024.

HAFS Ensemble	**2023**	**2024**
Members	20 + 1 control	30 + 1 control
Resolution	6km / 65 levels	4km / 81 levels
Microphysics	GFDL	Thompson
Ocean Model	HYCOM	MOM6/HYCOM
Stochastic Physics	SPPT, SKEB, SHUM	+SPP (ocean surface roughness)

Table 2: Comparison of the HAFS Ensemble configuration in 2023 (center) to 2024 (right).

The 31-member HAFS Ensemble performed quite well in 2024, and received positive feedback from NHC at the HFIP Annual Meeting. The season featured several challenging forecast cases, such as Hurricane Helene's 70-kt rapid intensification from 50 kt to 120 kt in 48 h beginning 25 September 2024 at 00 UTC. At this initialization time, approximately ⅓ of HAFS Ensemble members were explicitly predicting rapid intensification, which is a fairly strong signal for a phenomenon which climatologically only occurs approximately 5% of the time in any particular 24-h forecast period (Fig. 17).

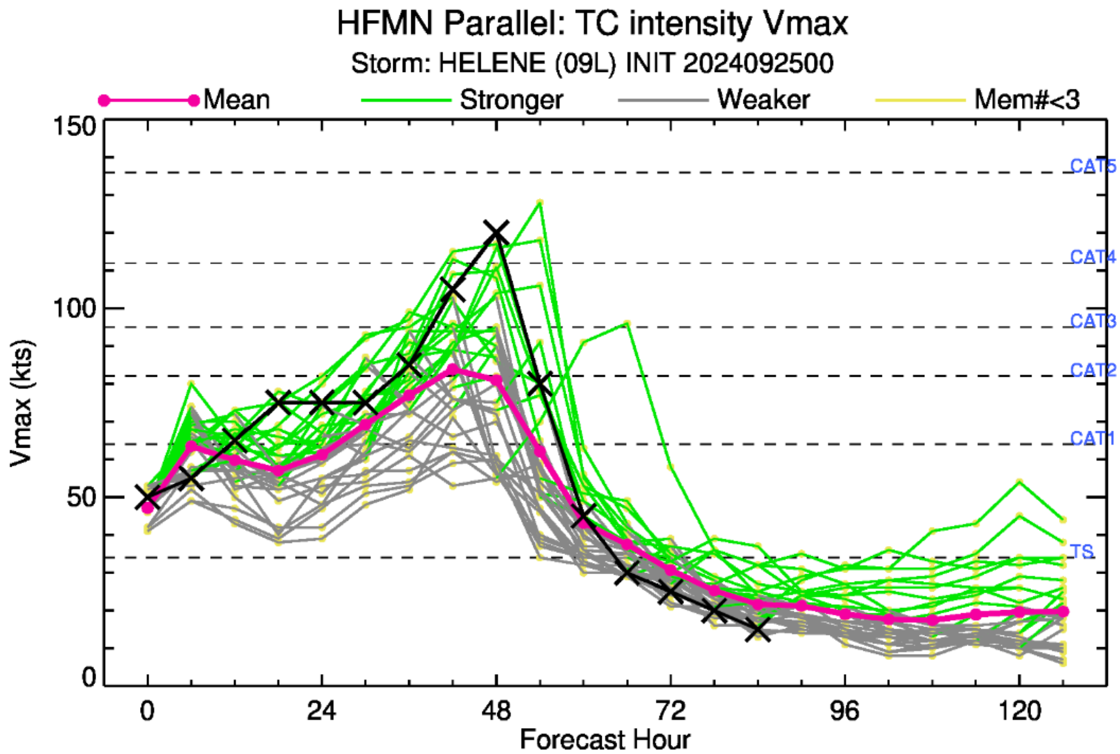


Figure 17: Intensity guidance from 0-120 h for Hurricane Helene, initialized 00 UTC 25 September 2024, from the 31-member HAFS Ensemble. The chart depicts the stronger 50th percentile (green), weaker 50th percentile (gray), the ensemble mean (magenta), and the verifying best track intensity (black).

HAFS Ensemble forecast guidance was also quite useful during the 2024 hurricane season in terms of depicting track forecast uncertainty. For example, shortly after the formation of Hurricane Milton at 12 UTC 06 October 2024, the HAFS Ensemble predicted the mean landfall location near Tampa Bay was a near-perfect forecast (Fig. 18). Across-track variability, or latitudinal variability in this case, around the time of landfall (day 4) was quite low, with all 31 members predicting a TC position either making landfall along the west coast of the FL peninsula, or just offshore. This is a remarkably high degree of track forecast agreement for a day 4 forecast. Note, however, that the day 4 and day 5 ellipses stretched significantly longitudinally, in the along-track direction. This indicates that, while the landfall location was relatively well known, there was significant uncertainty in the timing of landfall and eventual re-emergence over Atlantic waters off the east coast of FL, which varied in timing by as much as 36 h across all ensemble members.

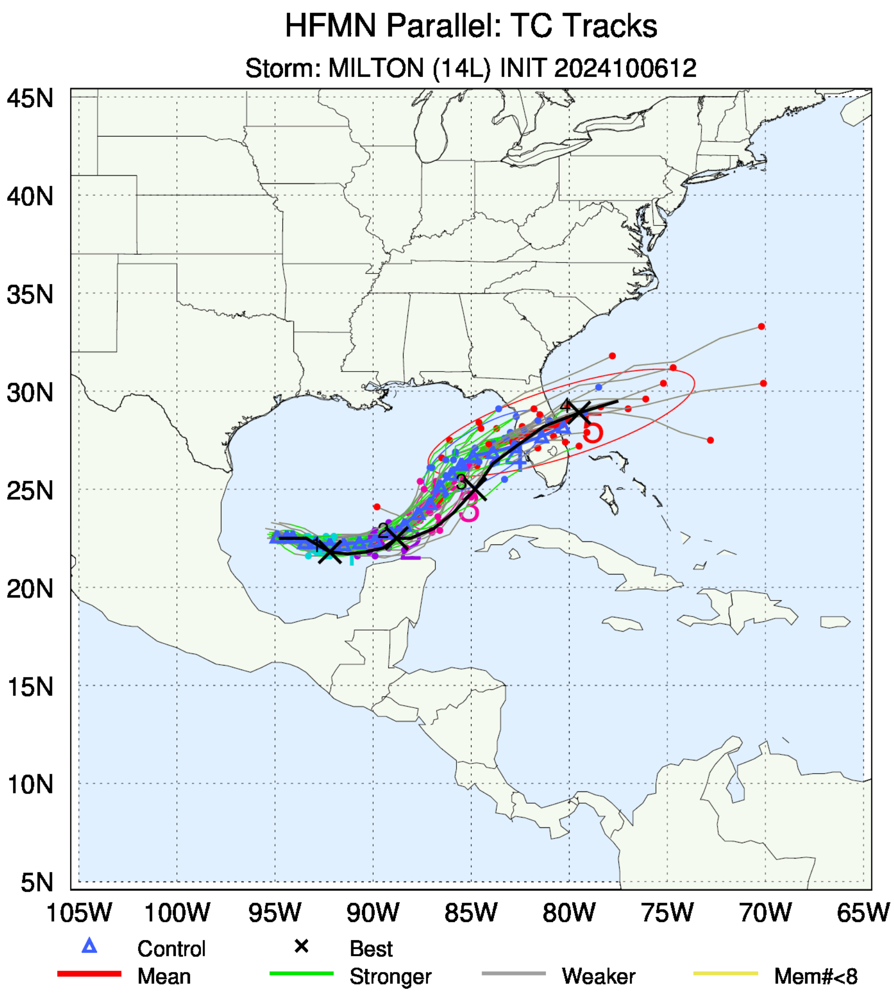


Figure 18: Five-day HAFS Ensemble track forecast for Hurricane Milton, initialized 12 UTC 06 October 2024. The strongest 50th percentile (green), weakest 50th percentile (gray), the ensemble mean (red), and the verifying best track (black) are shown. Numbers 1-5 indicate the ensemble mean center position at each day 1-5, and ellipses depict ⅔ of the ensemble position variance at each lead day. Ellipse major and minor axes accounting for the mean along-track and cross-track spread each day.

In addition to traditional track and intensity graphics, one of the HAFS Ensemble products which was well-received by NHC and WPC in 2024 was the probabilistic precipitation accumulation plots, which depict a map of ensemble mean accumulated precipitation in 24-hour time increments from the time of model initialization (t=0h) out to 120h, as well as total 0-120 h precipitation. An example from the 08 October 2024 12UTC forecast for Hurricane Milton is shown below (Fig. 19). Similar products which depict probabilities of precipitation exceeding various thresholds (e.g. P > 5" QPF) are also under development. Other high-resolution regional models such as the RRFS might be better suited for the prediction of other TC hazards, such as tornadoes. As HFIP expands in scope to encompass the full end-to-end value chain, we hope to include some basic verification for tornadoes in future reports.

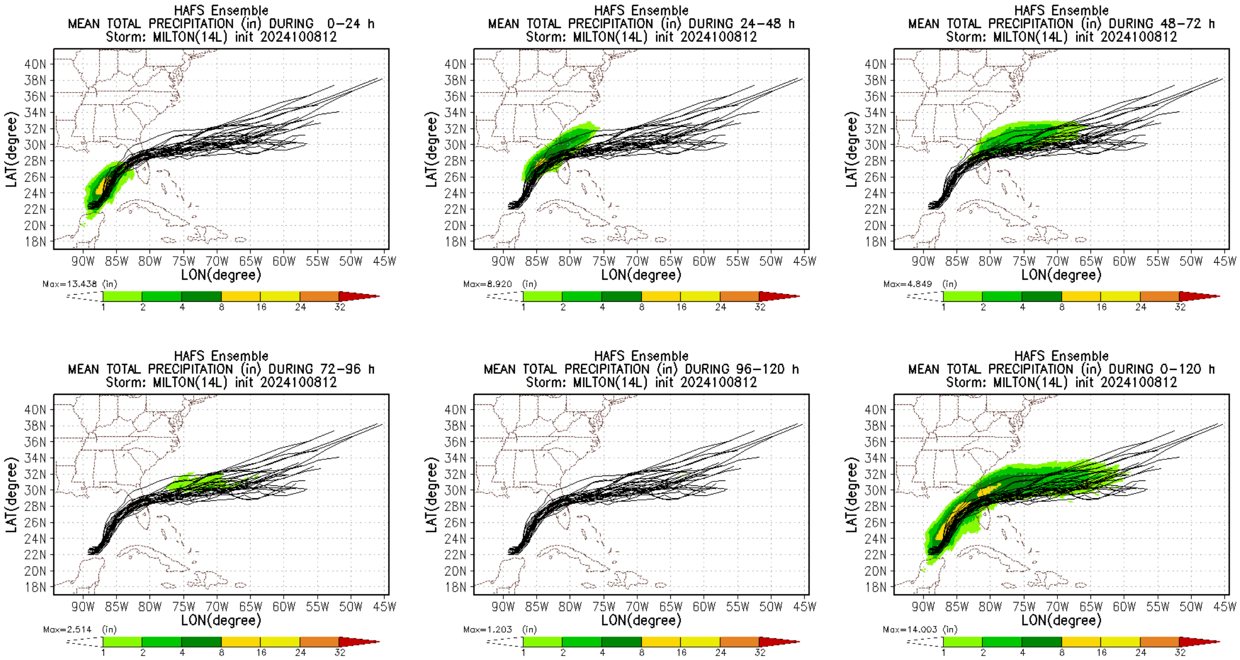


Figure 19: Ensemble mean precipitation (in) for Hurricane Milton from the HAFS Ensemble forecast initialized 12 UTC 08 October 2024, in 24-h increments from 0-120 h, as well as total 0-120 h precipitation. Track forecasts from the 31 individual ensemble members are shown as black lines.

Some major takeaways from the HAFS Ensemble in 2024 include the fact that the spread-error relationship is meaningfully improved compared to the version of 2023. The initial negative intensity bias from 2023 has also been largely removed by merging the operational HFSA vortices into basin domain fields. A major goal for the 2025 HAFS Ensemble will be to reduce the intensity error/bias for the first 3 days.

3. Community & Stakeholder Engagement: HFIP Annual Meeting 2024

The HFIP Annual Meeting 2024 was a hybrid meeting that took place in Miami, FL and online from November 12-15, 2024. This year's meeting was the largest to date, in terms of number of presenters and attendees, both in-person and online. Attendance included 96 in-person attendees and 184 unique online participants across the four days (a subset of the total attendance is shown in Fig. 20). The primary objective of the 2024 meeting was to discuss key HFIP strategies and present HAFS operational assessments from the v1 and v2 releases utilized in 2023 and 2024, as well as present and discuss early results and lessons learned from real-time developmental experiment results for future upgrades to HAFS. This information was used to inform and outline the work required to foster efficient pathways toward a world-leading, reliable, and skillful model guidance on TC track and intensity (including rapid intensification), storm size, genesis, storm surge, rainfall, and tornadoes associated with TCs and associated Socio-Economic impacts.

Additional accomplishments from the HFIP Annual Meeting 2024 included discussion of actionable plans to address draft 5- and 10-year goals and objectives of the HFIP Strategic Plan 2025. The goals discussed include focus on advancing forecasts and communication of all hazards from TCs; and incorporate risk communication research to develop more effective watch & warning, and probabilistic risk products with a focus on vulnerable communities and industries through the use of social, behavioral, and economic sciences. Additionally, rigorous discussion took place to address novel approaches for further enhancement of the HAFS role as the UFS Hurricane application while fostering even deeper integration of ideas and potential from other aspects of the larger UFS community both inside and outside of the TC realm.

Figure 20: Group photo on the afternoon of day 2 of the HFIP Annual Meeting, Wednesday 13 November 2024.

A number of forecast challenges were identified by NHC at this year's annual meeting, as highlighted here. In terms of tropical cyclogenesis, or "genesis" for short, predicting genesis is almost exclusively reliant on global model data. This is because most regional models are not initialized until there is a trackable disturbance or "Invest". One problem that arises due to reliance on global models for genesis is that cumulus is parameterized, which often contributes to model biases. NHC has found that biases are often both location dependent (e.g., western Caribbean Sea), and environment dependent (e.g., proximity of dry air).

NHC also identified a number of intensity and structure related issues. The shared microphysics parameterization (Thompson) between HAFS-A and HAFS-B in version 2.0 within the North Atlantic basin, while potentially leading to a slight reduction in mean error, on average, tends to produce excessive similarity in forecasts from the two HAFS variants. This can be problematic, as it can show forecasters too narrow a possible outcome of solutions. As noted elsewhere, overall HAFS seasonal intensity skill is lagging behind older hurricane-regional model guidance (HWRF/HMON). In terms of TC structure, when GFS develops TCs, it often is too fast to develop a symmetric vertically aligned vortex. HAFS models often initialize with unrealistic structures, including oddly shaped eyes / eyewalls, even for very intense TCs. This appears to be perhaps related to the DA and/or VI, and does not appear to be an issue in HWRF/HMON. There are also issues with the inner nest of HAFS centering on spurious convection instead of TC vortex center, which is more prominent in weaker TC or runs at the Invest stage.

Lastly, while we have made enormous strides in terms of track prediction since the advent of HFIP, a number of issues in forecasting TC track remain. Particular trends seen in 2024 include the GFS bias of being too fast and too far right with its track forecast, especially for strong hurricanes. The ECMWF has been almost universally too weak with TC intensity, which tends to cause its track to be too far south and west.

Feedback from WPC also provided some new perspective, particularly on the “all hazards” impacts from TCs, which includes deadly and extremely damaging freshwater flooding. WPC noted the importance of being able to first predict, and second communicate, where expected rainfall lands on the spectrum from ordinary to extraordinary (Fig. 21). The public and emergency management already understand that all TCs produce some rainfall, and many produce localized flooding. The challenge is being able to predict the relatively infrequent extreme rainfall producers, as well as which regions are most at risk from any particular storm. Hurricane Helene from the 2024 season was a particularly noteworthy example of the dangers of extreme rainfall due to a combination of a TC, orographic enhancement, and a predecessor rain event in the days prior to the TC which preconditioned the ground to already be saturated prior to the arrival of the main event.

Figure 21: The spectrum of precipitation risk associated with tropical cyclones, from ordinary to extraordinary (courtesy Mark Klein, WPC).

WPC has also done some verification work in terms of object-based verification of QPF. Similar to NHC's forecast, which combines track with intensity, the WPC forecast combines maximum amounts and locations of heaviest rainfall. One verification metric they are using for TCs is the displacement error in the location of the 2" rainfall contour, between their official forecast and the verifying rainfall map. Average errors at 24, 72, and 120 h appear in Table 3. WPC also verifies direction displacement. Predicted QPF for tropical cyclones is largely "in-step" with NHC official track. This would suggest a "slow" track bias causing the QPE centroid to be offset to the northeast. WPC has found a typical directional displacement of the observed QPE to the northeast of the forecast QPF in the majority of cases across all thresholds and forecast periods.

2016-2023 Displacement Error of 2" Rainfall Contour	
Lead Time	Avg. Error
24 hours	53 miles
72 hours	95 miles
120 hours	151 miles

Table 3: Average displacement error (mi) of WPC's 2" rainfall contour for landfalling TCs from 2016-2023 at 24 h, 72 h, and 120 h lead times.

While improving NWP for TC prediction has been the cornerstone of the HFIP program since its inception, observations and TC sampling have also been a critical component in terms of improving our fundamental understanding of TCs, improving physics parameterizations, for data assimilation, and for verification. Several presentations at the HFIP Annual Meeting 2024 highlighted how significantly we have advanced over the last several decades in terms of our ability to sample the inner core of the TC, the planetary boundary layer, as well as the upper ocean below the TC, with the accelerated advancement of unmanned aerial systems (UASs), uncrewed surface vehicles (USVs), gliders, drifters, floaters, etc, as part of NOAA's Advancing the Prediction of Hurricanes Experiment (APHEX) (Fig. 22a). UASs such as the Black Swift S0 and the Altius-600 can be deployed by the P-3 and continue flying, continually sampling the TC and reporting data in concert with the P-3, as was demonstrated with Hurricane Ernesto in 2024 (Fig. 22b), amongst other cases. Other more passive, but equally novel, systems such as streamsondes remain in the air more than twice as long as traditional dropsondes. As such they provided a much longer Lagrangian trajectory through the storm. Along these lines, HFIP continues to partner closely with the OAR/WPO Observations program to ensure that our programs work together to ensure that NOAA's model systems are positioned to take full advantage of new and emerging observational technologies.

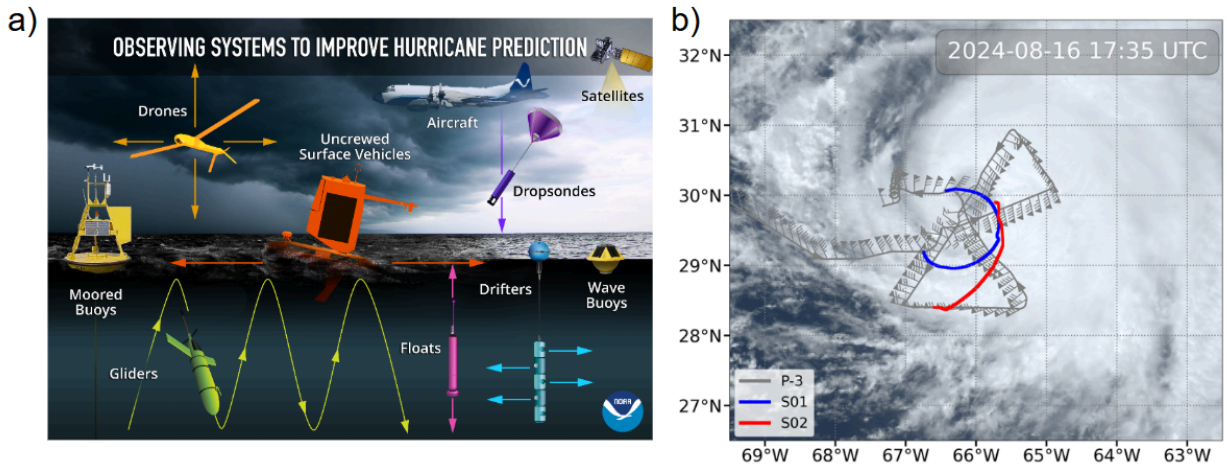


Figure 22: (a) Observational platforms used in 2024 to sample TCs by NOAA's APHEX experiment, as highlighted at the HFIP Annual Meeting (courtesy Jason Dunion); (b) Black Swift S0 deployment tracks (S01 & S02, blue & red respectively) into Hurricane Ernesto on 16 August 2024, circumnavigating the TC inner core and principal rainband in conjunction with the P-3 (grey).

This year's annual meeting also highlighted a recent shift toward AI/ML, in particular data driven AI weather prediction (AIWP) models. Preliminary results from a collaboration between Google DeepMind with their GraphCast system and EMC has produced a real-time GraphCast-GFS forecast system, using GDAS data for initial conditions. Verification of GraphCast-GFS from the 2024 season, along with three previous seasons of retrospective runs, provides encouraging early results (Fig. 23). Two parallel versions of the GraphCast-GFS outperform the GFS for TC track from 12-144 h, which would place the model in 1st or 2nd place for track as compared to any single deterministic solution (including ECMWF) at those lead times. However, as noted in their presentation, the GraphCast-GFS team encountered some challenges that remain in terms of intensity and structure prediction for AIWP. In particular, the TC is consistently weak biased and the vortex structure is too diffuse in AI models with respect to comparable resolution full physics NWP models.

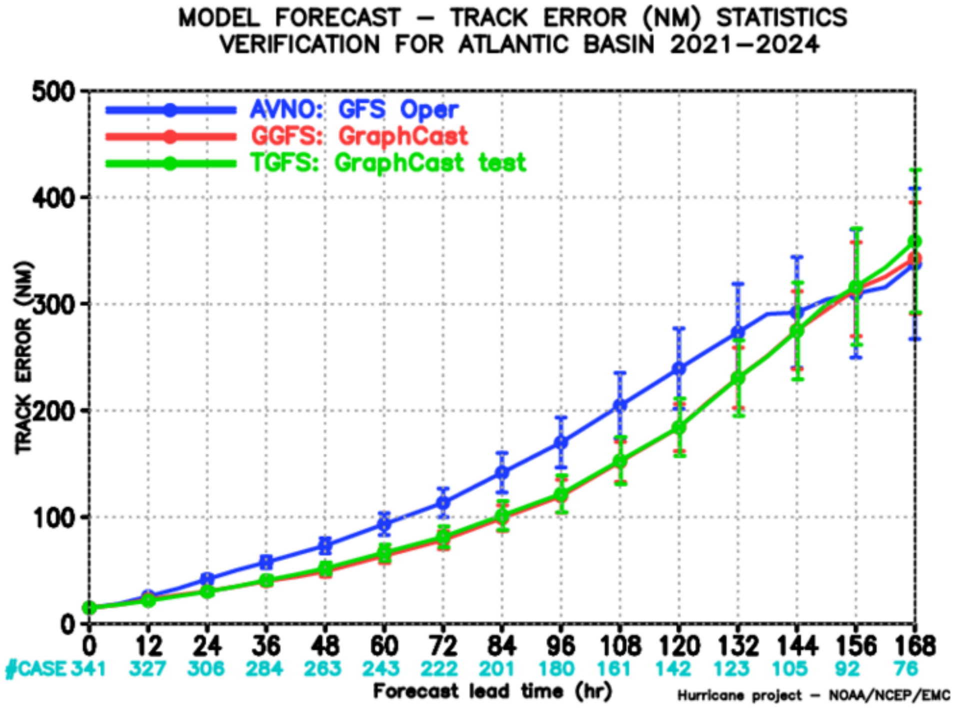


Figure 23: Track forecast error (n mi) from 0-168 h for the Atlantic basin from all retrospective and real-time runs from 2021-2024, comparing the operational GFS (blue) and two variations of GraphCast-GFS (red, green). The number of cases included in the sample at each forecast lead time is indicated along the x-axis in cyan.

The ECMWF team also presented some new results related to their AIFS model. The AIFS outperformed the deterministic ECMWF model for track at all lead times, which is a particularly impressive feat considering the ECMWF was already the top individual model for track in 2024. However, similar to the results for GraphCast-GFS, the AIFS lagged behind the deterministic system, and well behind mesoscale models such as HAFS, for predicting things like intensity (especially rapid intensification and rapid weakening), structure, and localized maxima in wind or precipitation.

4. Summary & Concluding Remarks

There were several key takeaways from HFIP in 2024, particularly related to future development of HAFS, increased forecast variance between HAFS-A and HAFS-B, a pathway for retiring HWRF and HMON, a path for the future operationalization of the HAFS Ensemble, what role AI will play in hurricane forecasting, the need for a HAFS reanalysis, better understanding and documenting the HFIP value chain, and expanding HFIP's reach through outreach and engagement.

For continued model development, verification, training of AI models, and for use in retrospective studies, there needs to be an organized effort to produce a HAFS reanalysis, ideally using the HAFS basin-scale configuration for East Pacific to Atlantic cross-basin continuity. However, this effort will require significant HPC and human resources. Similarly, there remains an open question as to what role AI models, in particular data-driven NWP, will play in the forecast process. In cooperation with Google DeepMind, the NWS is currently pursuing an AIWP “Graphcast GFS”. ECMWF’s AIFS model has also demonstrated impressive skill for track prediction, with plans in the works to improve upon intensity and structure prediction. How best to integrate AIWP guidance into the existing guidance suite remains under investigation.

One noteworthy challenge that remains is to maintain the NHC official forecast accuracy once HWRF/HMON are retired, which is anticipated for the 2026 hurricane season. While the HAFS suite now consistently outperforms HWRF/HMON for track and day 4-5 intensity on a seasonal basis, a skill gap versus the legacy models still remains for day 1-3 intensity. There will not be support to port the code of the legacy systems from WCOSS-2 to WCOSS-3, so there is a hard end date for these models once NCO makes the transition for the operational model platform. As such, a primary focus of the 2025 experimental model runs will be to improve the HAFS short-term intensity forecast. It was also noted that there is often insufficient forecast spread between HAFS-A and HAFS-B, so hurricane specialists will look to HWRF or HMON for alternate forecast scenarios. EMC has proposed a potential solution to this problem, in the event that HAFS-A and HAFS-B intensity skill continues to lag behind the legacy systems. EMC made the suggestion that they develop a HAFS-C variant, with greater independence of forecasts based upon a unique combination of physics and different resolution from HAFS-A and HAFS-B. There were some preliminary discussions that support from AOML/HRD would be provided for the development of HAFS-C. Since NHC typically uses a blended consensus approach, EMC suggested (with at least notional agreement from NHC) that if the blended consensus of HAFS-A, HAFS-B, and HAFS-C can outperform the blended consensus of HAFS-A, HAFS-B, HWRF, and HMON, then this would be acceptable criteria for retirement of the legacy systems.

While a prototype was not yet ready for testing in HREx 2024, development of the JEDI DA system and integration with HAFS remains a key priority for HAFS v3, and is one of the greatest short-term objectives of the revised Strategic Plan 2025-2035. Similar to the legacy hurricane models, the existing GSI DA system will no longer be supported in the near future. We anticipate a version of HAFS running with JEDI DA in HREx 2025, as well as an implementation plan mapping out an operational transition to HAFS with JEDI.

Another concern, particularly in regards to model development, is the ending of supplemental funding, with DRSA funding already having come to an end at the end of FY24, and FY26 being the final year of other supplemental funding. These funding sources account for a significant fraction of HFIP’s total budget for research and development (Fig. 24).

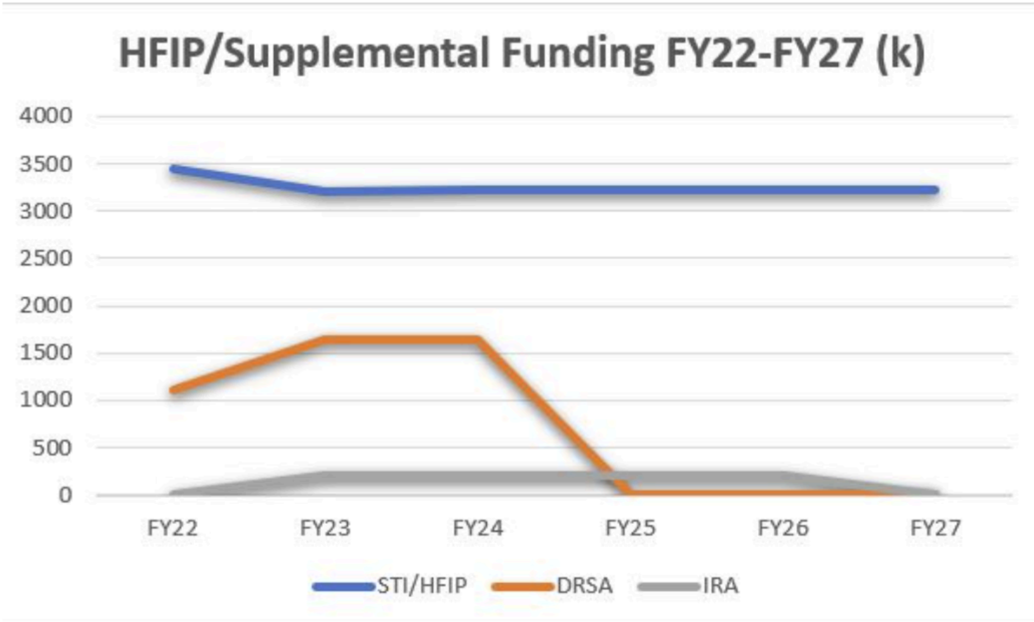


Figure 24: A gradual decrease in HFIP funding over the years, including an end to supplemental funding sources DRSA and IRA.

HFIP stakeholders voiced their concerns in 2024 that, while the research-to-operations (R2O) pipeline is working, there are areas in which efficiency can be optimized. We need to begin folding in transition details at the proposal stage, including resources required, to allow for sufficient planning between HFIP partners and adjusting the project governance as needed. Along these lines, it was apparent in 2024 that HFIP needs to think more about the end-to-end value chain; what do the public most need to know in order to make decisions? Can we get to the point where we can say: “there is a 60% chance of a 10-day power outage”? HFIP must continue to support outreach and engagement with students and younger scientists via colloquiums such as the HFIP/NCAS-M HAFS training colloquium from May 2024. These students and young scientists will be the future model developers, program managers, and risk communicators that continue the legacy of HFIP. Finally, for those who issue forecasts and communicate with the end users (NHC, WPC, local WFOs, etc), it is of paramount importance to know and target your audience. The same messaging will not work for all audiences, because all audiences will not respond the same way.

5. List of HFIP Publications in 2024

Journals and Periodicals

Alaka Jr., G,. J. A. Sippel, Z. Zhang, H.-S. Kim, F. Marks, V. Tallapragada, A. Mehra, X. Zhang, A. Poyer, and

S. G. Gopalakrishnan, 2024: Lifetime performance of the operational Hurricane Weather Research and Forecasting model (HWRF) for North Atlantic tropical cyclones. *Bull. Amer. Meteor. Soc.,* **105,** E932–E961, https://doi.org/10.1175/BAMS-D-23-0139.1.

Alvey, G., Alaka, G., Gramer, L., and A. Hazelton, 2024: Evaluation of Hurricane Analysis and Forecast System (HAFS) Forecast Error Statistics Stratified by Internal Structure and Environmental Metrics, *Wea. Forecasting*, **40**, 131-147, https://doi.org/10.1175/WAF-D-24-0030.1.

Aristizábal Vargas, M. F., H.-S. Kim, M. L. Hénaff, T. Miles, S. Glenn, and G. Goni, 2024: Evaluation of the ocean component on different coupled hurricane forecasting models using upper-ocean metrics relevant to air-sea heat fluxes during Hurricane Dorian (2019). *Front. Earth Sci.,* **12,** 1342390, https://doi.org/10.3389/feart.2024.1342390.

Bower, E., Reed, K. A, Alaka, Jr., G. A., and A. T. Hazelton, 2024: Verification of Operational Forecast Models in Cases of Extratropical Transition of North Atlantic Hurricanes, *Wea. Forecasting*, **39**, 1695–1714, https://doi.org/10.1175/WAF-D-24-0011.1.

Gramer, L. J., J. Steffen, M. Aristizabal, and H.-S. Kim, 2024: The impact of coupling a dynamic ocean in the Hurricane Analysis and Forecast System. *Front. Earth Sci.,* **12,** 1418016, https://doi.org/10.3389/feart.2024.1418016.

Hazelton, A. T., Chen., X., Alaka, G. J.. Jr., Alvey, G. R. III, and S. Gopalakrishnan, 2024: Sensitivity of HAFS-B Tropical Cyclone Forecasts to Planetary Boundary Layer and Microphysics Parameterizations, *Wea. Forecasting*, **39**, 655–678, https://doi.org/10.1175/WAF-D-23-0124.1.

Kim, H.-S., B. Liu, B. Thomas, D. Rosen, W. Wang, A. Hazelton, Z. Zhang, X. Zhang, and A. Mehra, 2024: Ocean component of the first operational version of Hurricane Analysis and Forecast System: Evaluation of HYbrid Coordinate Ocean Model and hurricane feedback forecasts. *Front. Earth Sci.,* **12,** 1399409, https://doi.org/10.3389/feart.2024.1399409.

Munsi, A., A. P. Kesarkar, J. N. Bhate, and V. Tallapragada, 2024: Helicity: A Possible Indicator of Negative Feedback Initiation of Tropical Cyclone–Ocean Interaction. *Earth and Space Science*, **11**, e2023EA003211, https://doi.org/10.1029/2023EA003211.

Munsi, A., Kesarkar, A.P., Bhate, J.N., and Tallapragada, V., 2024: Helicity: A Possible Indicator of Negative Feedback Initiation of Tropical Cyclone–Ocean Interaction. *Earth and Space Science*, **11**, e2023EA003211. https://doi.org/10.1029/2023EA003211

Peng, J., Z. Zhang, W. Wang, R. Panda, B. Liu, Y. Weng, A. Mehra, V. Tallapragada, X. Zhang, S. Gopalakrishnan, W. Komaromi, J. Anderson and A. Poyer, 2024: HAFS Ensemble Forecast in AWS. *Front. Earth Sci.,* **12,** 1396612, https://doi.org/10.3389/feart.2024.1396612.

Shin, J., Z. Zhang, B. Liu, Y. Weng, Q. Liu, A. Mehra, and V. Tallapragada, 2024: The implementation of cloud and vertical velocity relocation/cycling system in the vortex initialization of the HAFS. *Atmosphere,* **15,** 1006, https://doi.org/10.3390/atmos15081006.

Wang, W., J. Han, J. Shin, X. Chen, A. Hazelton, L. Zhu, H.-S. Kim, X. Li, B. Liu, Q. Liu, J. Steffen, R. Sun, W. Zheng, Z. Zhang, and F. Yang, 2024: Physics schemes in the first version of NCEP operational hurricane analysis and forecast system (HAFS). *Front. Earth Sci.,* **12,** 1379069, https://doi.org/10.3389/feart.2024.1379069.

Publications in Print

Xie, L. and B. Liu, 2024: Marine meteorology. *Encyclopedia of Atmospheric Sciences, 3nd Edition,* Elsevier Ltd., accepted.

Publications in Review

Chen, J.-H., T. Marchok, M. Bender, K. Gao, S. Gopalakrishnan, L. Harris, A. Hazelton, B. Liu, A. Mehra, M. Morin, F. Yang, X. Zhang, Z. Zhang, and L. Zhou, 2024: Closing the gap - advances in US models' hurricane predictions. *Bull. Amer. Meteor. Soc.,* in review.

Technical Reports:

Wang, W., Y. Weng, C. Wang, J. Peng, B. Liu, and Z. Zhang, 2024, An idealized hurricane analysis and forecast system (I-HAFS), WMO blue Book, Section 9.

Worthen, D., J. Wang, R. Montuoro, D. Heinzeller, B. Li, G. Theurich, U. Turuncoglu, D. Rosen, D. Jovic, B. Curtis, R. Mahajan, H. Lei, A. Richert, A. Chawla, J. Wang, J. Meixner, A. Abdolali, M. Masarik, L. Pan, M. Barlage, B. Liu, M. Vertenstein, T. Craig, R. Benson, T. Robinson, T. Clune, W. Jiang, N. Barton, G. Vandenberghe, M. Potts, J. Kim, N. Perlin, C. Book, L. Bernardet, F. Yang, S. Sun, H.-S. Kim, B. Baker, J. Huang, C.-H. Jeon, and I. Stajner, 2024: Coupling infrastructure capability in UFS weather model. *Office Note (National Centers for Environmental Prediction (U.S.)) 519,* 115pp, https://doi.org/10.25923/dvv2-3g03.

Appendix A: Table of Acronyms

3DIAU	3-Dimensional Incremental Analysis Update (3DIAU)
4DEnVar	4-Dimensional Ensemble Variance-based data assimilation
AFS	Analyze, Forecast and Support office
AI	Artificial Intelligence
AIWP	Artificial Intelligence Weather Prediction
AMV	Atmospheric Motion Vector
AOML	Atlantic Oceanographic and Meteorological Laboratory
AOR	Area of Responsibility
AWIPS	Advanced Weather Interactive Processing System
AWS	Amazon Web Services
C3	Community Convective Cloud
CIMAS	Cooperative Institute For Marine And Atmospheric Studies
CIRA	Cooperative Institute for Research in the Atmosphere

CLP5	5-day Climatology and Persistence Track Forecast
COAMPS-TC	Coupled Ocean-Atmosphere Mesoscale Prediction System for Tropical Cyclones
CONUS	Continental United States
CPHC	Central Pacific Hurricane Center
DA	Data Assimilation
DESI	Dynamic Ensemble-based Scenarios for IDSS
DSHF	Decay SHIFOR Model Intensity Forecast
DSRC	David Skaggs Research Center
DTC	Developmental Testbed Center
ECMWF	European Centre for Medium-Range Weather Forecasts
EDMF	Eddy-Diffusivity/Mass-Flux
EMC	Environmental Modeling Center
EPAC	East Pacific
EPIC	Earth Prediction Innovation Center
ePBL	energetics-based Planetary Boundary Layer
ERC	Eyewall Replacement Cycle
FAR	False Alarm Ratio
FY	Fiscal Year
GEFS	Global Ensemble Forecast System
GFDL	Geophysical Fluid Dynamics Laboratory
GFS	Global Forecast System
GOES	Geostationary Operational Environmental Satellites
GSI	Gridpoint Statistical Interpolation
GSL	Global Systems Laboratory
HAFS	Hurricane Analysis and Forecast System
HEOB	HFIP Executive Oversight Board
HFIP	Hurricane Forecast Improvement Program
HMON	Hurricanes in a Multi-scale Ocean-coupled Non-hydrostatic model
HOT	Hurricane Ocean Testbed
HPC	High Performance Computing
HRD	Hurricane Research Division
HREx	HFIP Real-time Experiments
HWRF	Hurricane Weather Research and Forecast model
HYCOM	Hybrid Coordinate Ocean Model
IDSS	Impact-based Decision Support Services

IOC	Initial Operating Capability
JEDI	Joint Effort for Data assimilation Integration
JTWC	Joint Typhoon Warning Center
ML	Machine Learning
MOM6	Modular Ocean Model v6
MYNN	Mellor–Yamada–Nakanishi–Niino
NATL	North Atlantic
NCEP	National Centers for Environmental Prediction
NESDIS	National Environmental Satellite, Data, and Information Service
NHC	National Hurricane Center
NOAA	National Oceanic and Atmospheric Administration
NOS	National Ocean Service
NWS	National Weather Service
OAR	Oceanic and Atmospheric Research
OCD5	Operational CLP5 and DSHF Blended Intensity Forecast
OSTI	Office of Science and Technology Integration
PBL	Planetary Boundary Layer
PDF	Probability Density Function
POD	Probability of Detection
R&D	Research and Development
R2O	Research-to-Operations
R34	Radius of 34-kt wind
RDHPC	Research and Development High Performance Computing
RI	Rapid Intensification
RL	Readiness Level
RMW	Radius of Maximum Wind
RRTMG	Rapid Radiative Transfer Model
RTOFS	Real-Time Ocean Forecast System
SA-SAS	Scale-Aware Simplified Arakawa-Schubert
SAS	Simplified Arakawa-Schubert
SBES	Social Behavioral and Economic Science
SHiELD	System for High-resolution prediction on Earth-to-Local Domains
SHIFOR	Statistical Hurricane Intensity Forecast
SHUM	Stochastic Humidity perturbations
SKEB	Stochastic Kinetic Energy Backscatter
SPC	Storm Prediction Center

SPPT	Stochastically Perturbed Parametrization Tendencies
STAR	Center for Satellite Applications and Research
T-SHiELD	Tropical System for High-resolution prediction on Earth-to-Local Domains
TC	Tropical Cyclone
TCANE	Tropical Cyclone Artificial Neural Network Error
TKE	Turbulent Kinetic Energy
UFS	Unified Forecast System
UQ	Uncertainty Quantification
UTC	Coordinated Universal Time
VI	Vortex Initialization
WFO	Weather Forecast Office
WPAC	West Pacific
WPC	Weather Prediction Center
WSP	Wind Speed Probability
WTCM	Windspeed Tropical Cyclone Model
WW3	Wavewatch III