The HAFS Ensemble in Real-time on the Cloud (HERC) 2023 - Accomplishments and Lessons Learned

William Komaromi[1], Aaron Poyer[2], Zhan Zhang[3], Jiayi Peng[4], Rajendra Panda[5], Xuejin Zhang[6], Sundararaman Gopalakrishnan[6], Ghassan Alaka[6], Jason Anderson[1], Kevin Garrett[2], Frank Indiviglio[7], Unni Kirandumkara[8], Gonzalo Lassally[7], Matt Long[9], Avichal Mehra[3], Radhika Subramanian[7], Vijay Tallapragada[3]

October 2023

1. NOAA/NWS/OSTI-Modeling & IM Systems Group
2. NOAA/NWS/OSTI-Modeling
3. NOAA/NCEP/EMC
4. AXIOM & NOAA/NCEP/EMC
5. Yayworks Inc. & NOAA/NCEP/EMC
6. NOAA/AOML/HRD
7. NOAA/OCIO/HPCC
8. GDIT Inc.
9. Parallel Works

Overview

The Hurricane Analysis and Forecast System (HAFS) Ensemble in Real-time on the Cloud (HERC) experiment of 2023 was a highly successful real-time demonstration of the feasibility of the National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS) running a state-of-the-science hurricane modeling system on the cloud. The HERC experiment, supported by Disaster Relief Supplemental Appropriations (DRSA) Hurricane Project 2 (HURR2) supplemental funding of the Hurricane Forecast Improvement Program (HFIP), advanced the state of ensemble science and development at the National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC). This was realized through real-time Amazon Web Service (AWS) cloud allocation provided by the NOAA Office of the Chief Information Officer (OCIO) High Performance Computing & Communications (HPCC). HERC ran from August 1 through November 30, 2023, coinciding with peak activity and the latter half of the 2023 Atlantic hurricane season. HERC was run on AWS using the Parallel Works middleware platform. Running a 21-member ensemble of NOAA's novel HAFS model, the deterministic version of which became operational in 2023, real-time probabilistic forecasts for Atlantic tropical cyclones were produced 4x per day.

In this document, we will discuss our experimental design for the HERC experiment, including the model configuration and cloud computing setup. We will then provide a few example forecasts and highlight positive feedback we have received from stakeholders. Lastly, we will discuss lessons learned and propose future directions for HERC moving into the 2024 hurricane season.

Model Configuration

The Hurricane Analysis and Forecast System (HAFS) is a newly developed hurricane modeling system featuring the Finite-Volume Cubed-Sphere Dynamical Core (FV3) core from NOAA's Geophysical Fluid Dynamics Laboratory (GFDL), as part of the Unified Forecast System (UFS). Two configurations of deterministic HAFS became operational at the NOAA National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC) in 2023, HAFS-A and HAFS-B. The HERC configuration more closely matches the HAFS-A model configuration, although there are differences, primarily related to challenges in setting up cycled data assimilation (DA) on the cloud. A single domain spans the North Atlantic basin, approximately 5°S-58°N, 110-10°W, with a uniform 6-km horizontal resolution grid and 66 vertical levels. Due to a number of technical hurdles and time constraints in getting the ensemble up and running for peak seasonal activity, vortex initialization (VI) and DA are disabled for HERC in its debut season. For comparison, VI and vortex-scale DA are used for tropical cyclones (TCs) of initial intensity greater than 50 kt for operational HAFS run on the on-premise NOAA Research and Development High Performance Computing (RDHPC) systems. Land surface conditions are provided by the Noah Land-Surface Model (LMS) with Visible Infrared Imaging Radiometer Suite (VIIRS) vegetation, and the ocean is two-way coupled to the HYbrid Coordinate Ocean Model (HYCOM). The Global Forecast System (GFS) surface layer is employed, with TC-specific sea surface roughness adopted from the Hurricane Weather Research Forecast (HWRF) model. Additional physics options used for HERC include the Scale-Aware Turbulent Kinetic Energy Eddy-Diffusivity/Mass-Flux (Sa-TKE-EDMF) boundary layer with TC-related calibration and mixing length adjustments, GFDL single-moment microphysics, Rapid Radiative Transfer Model (RRTMG) radiation, the scale-aware Simplified Arakawa–Schubert (SAS) parameterization for deep and shallow cumulus, and Unified Gravity Wave Physics version 1 (uGWPv1) gravity wave drag. The ensemble is run four times per day, at 00, 06, 12, and 18 UTC.

For HERC, the ensemble is running with an unperturbed HAFS-A control member, plus 20 perturbed HAFS-A members, generating 21 ensemble members in total. Initial conditions, lateral boundary conditions, and model physics are all perturbed in order to generate diversity and appropriate forecast spread amongst ensemble members. Initial condition and boundary condition perturbations are provided by the Global Ensemble Forecast System (GEFS) operational system at 0.5x0.5 degree resolution. Stochastically perturbed physics tendencies (SPPT) represents uncertainties in physical parameterizations, and multiplicative noise modifies the total parameterized tendency. Stochastic kinetic energy backscatter (SKEB) counteracts excessive energy dissipation from numerical diffusion and interpolation, mountain and gravity wave drag, and deep convection. The stream function is randomly perturbed to represent upscale kinetic energy transfer. Stochastically perturbed planetary boundary layer (PBL) humidity (SHUM) represents variability in the sub-grid humidity field, similar to SPPT, but directly modifies low-level humidity field instead of heating and moisture tendencies.

Cloud Computing Setup

Cloud HPC resources for the HERC experiment were provided by Amazon Web Service (AWS) using the Parallel Works middleware platform. There were 3 challenges to be overcome for a successful implementation of HERC. They were: (1) estimation of cloud computing cost for a real-time simulation of 21-member ensemble forecast that needed to be done 4 times a day over the hurricane season lasting 3 to 4 months, (2) managing the overall cloud computing cost within project budget, and (3) porting and migration of the HAFS workflow to AWS as well as testing of 21-member ensemble forecast workflow on the cloud.

Performance information for a single member HAFS on a NOAA R&D HPC system was used for a rough estimate of required cloud resources and runtimes. To meet the first challenge, 168 hpc6a.48xlarge instances (nodes) are used. Each forecast member requires 7 nodes per single forecast job (per ensemble member) or a total of 147 nodes for all 21 members. The remaining 21 nodes are used to run other (e.g. pre and post processing) jobs in the workflow. This configuration was implemented using Amazon Reserved Instances (RI) starting on August 1, 2023.

Amazon Reserved Instances provide a significant discount relative to On-Demand Instances. RIs can be used for capacity reservation over a duration of time; however, the project is charged whether RIs are used or not over the entire duration. The second challenge was met by deciding to go with RIs which led to the third challenge of porting and migrating HAFS workflow to AWS but also minimizing testing of the full (21-member) ensemble workflow runs on the cloud configuration with RIs. The objective was to have the full ensemble workflow completely tested and ready to go by August 15 when climatological peak Atlantic hurricane activity begins. We allocated a “spin-up” period of 6 weeks during which HAFS ensemble workflow could be migrated and tested on AWS.

Fortunately, by having a 6-week “spin-up” period, the team was able to get the model running reliably in real-time with virtually no down time. The spin-up period ran from July 3 through August 11, which gave the team time to install and configure HAFS on AWS, ensure that the necessary static libraries were accessible, ensure that the necessary modules were available, and that the pipeline for data archiving and real-time dissemination were all working. The spin-up period was divided into 2 phases. During the first phase, as part of the migration process only 1-3 member ensembles were tested on AWS using the On-Demand Instances. Only 7 to 21 instances were used to keep the costs low. During the second phase of the spin-up period, the full ensemble was tested on the RI-based cloud configuration. In the August - September period, there were only two forecasts in which one of the 21 ensemble members failed, versus 5122 completed forecasts, for a success rate of 99.96%. There were then two missed forecasts at the start of the new fiscal year on October 1 related to transitioning account resources to start the new fiscal year. These forecasts were subsequently re-run successfully.

In terms of storage allocation needs, the HERC project utilized 48 TB of a persistent Lustre file system. Approximately 4.5 TB are needed to hold the HAFS code, static databases, input files, and temporary space needed for the model to run. The data from each cycle of the 21-member ensemble requires 21 TB disk space, so the 48 TB was sufficient to hold two cycles before transferring the data to a more permanent storage S3 bucket. While each new cycle is running, the previous completed cycle was in the process of transferring to the S3 bucket, as well as a limited set of files being transferred to the Research and Development High Performance Computing Systems (RDHPCS) on-premise machines for generating graphics for the EMC and

NOAA Oceanic and Atmospheric Research (OAR) Atlantic Oceanographic and Meteorological Laboratory (AOML) Hurricane Research Division (HRD) experimental HFIP run websites.

Example Forecasts & Preliminary Verification

Recent studies have shown that high-resolution ensembles for tropical cyclones have the potential to outperform a single deterministic forecast for both track and intensity. Figure 1a shows an example track forecast for Hurricane Idalia initialized 0600 UTC 27 August 2023. In this forecast, the unperturbed control member was too far west with Idalia's landfall in the FL panhandle, while the ensemble mean forecast was almost dead-on compared to the verifying "best track". The degree of spread amongst ensemble members also suggested a moderate to low degree of forecast uncertainty

In contrast to the Idalia forecast, Figure 1b depicts the ensemble track forecast for Tropical Storm Philippe from 0600 UTC 1 October 2023. The very large spread in the ensemble indicates high uncertainty. In this case, the verifying track occurred west of all 21 members at day 3, and west of 20 out of 21 members at day 5. In this case, the models particularly struggled to handle the binary interaction between Philippe and Tropical Storm Rina. Very large track uncertainty was not limited to the HAFS ensemble; the GEFS and European Center for Medium-range Weather Forecasting (ECMWF) ensembles also depicted very large spread for track at this time.

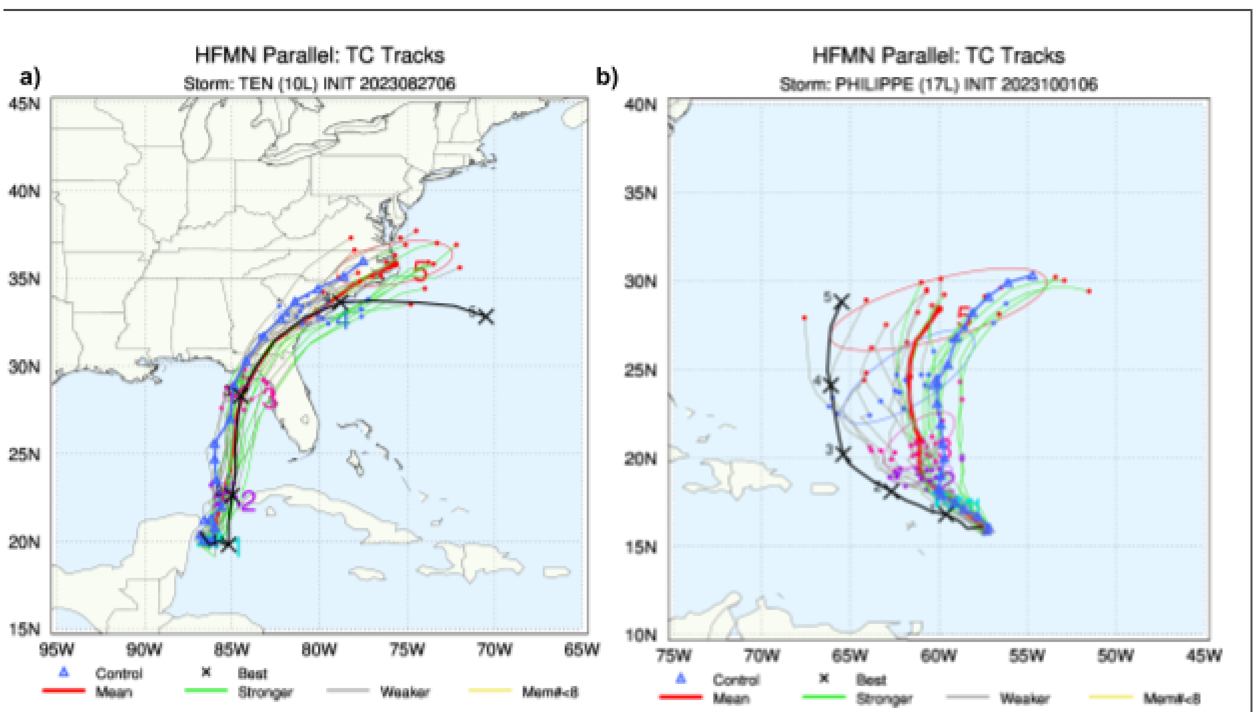

Figure 1: *HAFS Ensemble track forecast for (a) Hurricane Idalia initialized 0600 UTC 27 August 2023, and (b) Tropical Storm Philippe initialized 0600 UTC 01 October 2023. The unperturbed control member is blue, the ensemble mean is red, and the verifying track is black.*

An example intensity forecast for Hurricane Idalia initialized 1800 27 August 2023 is shown in Figure 2. In this forecast, the ensemble mean peak intensity prior to landfall is 25-kt too weak. However, 5 of the 21 members depict rapid intensification during the 36-60 h timeframe with a peak intensity of Category 3 or greater, suggesting a ~24% chance that the storm would be significantly stronger than the mean. This demonstrates the usefulness of an ensemble when making tropical cyclone intensity predictions. Given this forecast, the forecaster could emphasize the large spread in the intensity forecast and communicate the risk of a stronger than predicted hurricane.

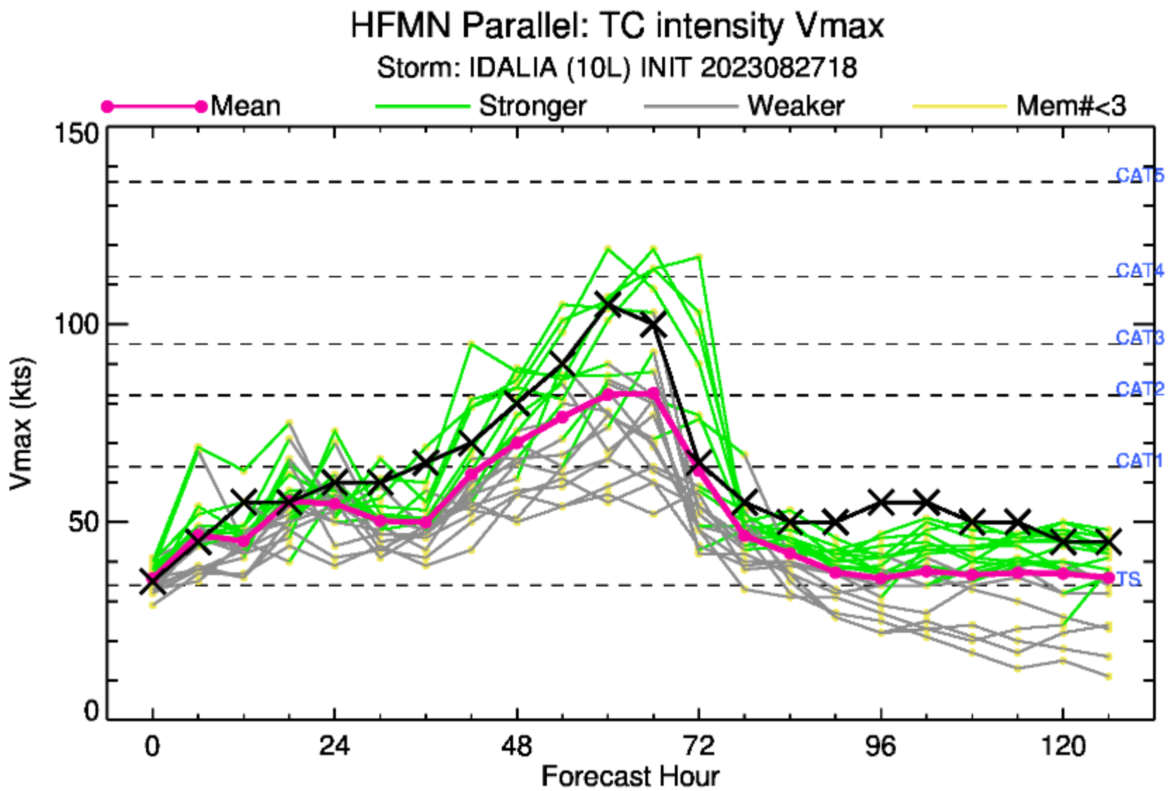

Figure 2: *HAFS Ensemble intensity forecast for Hurricane Idalia initialized 1800 UTC 27 August 2023. The forecasts are divided into two clusters, one with the stronger members (green) and one with the weaker members (gray). The ensemble mean is magenta, and the verifying intensity is black.*

Season-to-date error statistics since the start of the HERC real-time forecast period, from 01 August 2023 through 10 October 2023, for track and intensity have been computed in order to assess the overall performance of the system (Figure 3). Overall the HAFS Ensemble in Real-time on the Cloud system performed competitively for track, with errors 5-10% greater than the operational HAFS-A and GEFS mean, on average. Somewhat surprisingly, the unperturbed control member was associated with lower mean error than the ensemble mean. However, the sample only includes part of a single season within a single basin, and is < 25% the size of a full retrospective sample used to show statistical significance. We expect that the ensemble mean will produce lower error than the unperturbed control over a larger sample, as has been demonstrated on numerous ensemble systems in the literature. It is also possible that leveraging GEFS ensemble perturbations, as was done for HERC this season, does not perturb the optimum modes of error growth to produce track deviations of equal likelihood. A future perturbation technique such as a cycled hybrid Ensemble Kalman Filter / Variational (EnKF-Var), which projects more directly upon error growth in HAFS itself, may be more successful. The team will remain cognizant of this in future testing with DA.

It should also be emphasized that the operational HAFS-A and GEFS are a high bar for success, and amongst the most skillful sources of track guidance available to forecasters. Given the limitations of the debut HERC system, such as 6-km horizontal resolution versus 2-km resolution for deterministic HAFS, and no data assimilation or vortex initialization, track error close to that of HAFS-A and GEFS is considered a success.

The HERC ensemble performed more competitively for intensity, with the control member and ensemble mean both associated with lower mean error than the GEFS from 0-72 h. The HAFS ensemble mean forecast intensity also outperformed the operational HAFS-A from 30-96 h, which is an impressive accomplishment. Given the competitive statistics of this demonstration system for intensity at numerous lead times, we expect future versions of the HAFS ensemble with 2-km nesting, DA/VI, etc., to be amongst the top sources of intensity guidance at most, if not all, lead times.

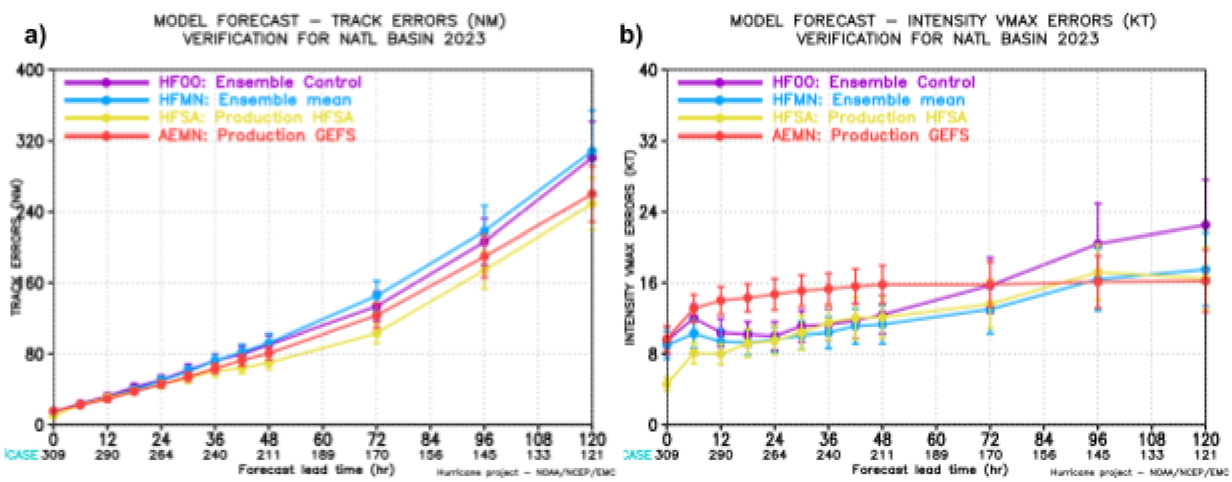

Figure 3: *Season-to-date (a) track (km), and (b) intensity (kt) forecast errors as a function of lead time from 01 August 2023 through 10 October 2023 for the unperturbed HAFS ensemble control member (magenta), the HAFS ensemble mean (blue), the operational HAFS-A (yellow), and the operational GEFS ensemble mean (red).*

Summary & Future Directions

The HERC experiment demonstrated the ability to transition NOAA software or models from on-premise systems to running on the cloud on a timescale of several weeks, given the right team working together with the necessary expertise. In the August - September period, there were only two forecasts in which one of the 21 ensemble members failed, versus 5122 completed forecasts, for a success rate of 99.96%. The forecasts were timely enough to be useful for National Hurricane Center (NHC) forecasters, and were posted in real-time to EMC's experimental hurricane forecast webpage (https://www.emc.ncep.noaa.gov/HAFS/HAFSEPS/) with very little latency. The experiment also identified challenges associated with transitioning to the cloud, such as difficulty in staging the necessary data to run the data assimilation and vortex initialization, which was turned off for this experiment for simplicity. Challenges in running licensed graphics plotting software on the cloud also necessitated transferring files to on-premise NOAA HPC machines in order to generate the graphics for EMC's experimental hurricane forecast webpages. The HERC forecasts are archived on the NWS's AWS S3 storage bucket for further analysis in the off-season.

Due to the resounding success of this experiment, we plan to run HERC again in the 2024 Atlantic hurricane season using an updated version of the model and including DA and VI in the cloud workflow. In future seasons, we would like to introduce a second HAFS domain that includes the eastern and central North Pacific, since these regions are also within NHC's area of responsibility. The team has also recently overcome a technical hurdle which had previously prevented us from running with multiple storm-following nests. Assuming that the team can ensure that sufficient compute resources are provided, we will introduce multiple storm-following 2-km resolution nests for next season.