The Developmental Testbed Center

HWRF 2011 Baseline Test Report

Point of Contact: Ligia Bernardet (ligia.bernardet@noaa.gov)

May 2011

Contributors:

The DTC HWRF 2011 Baseline Test was conducted by the team of the Hurricane Task of the DTC, composed by Shaowu Bao, Ligia Bernardet, Timothy Brown, Laurie Carson, Christopher Harrop, and Donald Stark. This test was designed in collaboration with Vijay Tallapragada of NOAA/NCEP/EMC, who provided the HR20 tracks for verification and comparison against HNR2.

All acronyms are defined in Appendix C.

Index

1. Executive Summary

- The DTC conducted its first extensive test of a HWRF configuration, demonstrating that the development of a testing infrastructure functionally equivalent to NOAA EMC's is complete.
- Over 1000 HWRF runs for the Eastern North Pacific and Atlantic basins, for the 2008, 2009, and 2010 seasons were conducted in order to produce a robust Reference Configuration of the community code (HWRF PS:85.98.98.88.88.2.4 nicknamed HNR2).
- Track errors for HNR2 increase linearly with time from near zero at initialization time to 280 nm at the 5-day forecast.
- A negative intensity bias is noted for HNR2 at all lead times, with a marked increase in errors in the first 6-h of the forecast, suggesting a problem with the initialization. More recent results, not presented in this report, indicate that this problem has been mitigated in more recent HWRF configurations.
- Absolute intensity errors increase sharply in the first 6-h of forecast and then grow slowly out to 3-days, after which they remain virtually unchanged.
- The forecast storm size is larger than the observed one for the 34-, 50-, and 64-kt wind radii, with the worst errors occurring for the 34-kt radii.
- A comparison between HNR2 and HR20 (a similar HWRF test conducted by EMC) was conducted for the purposes of assessing how similar the forecasts produced with the community code were to forecasts produced with a similar configuration from the EMC code repository. An exact match was not expected due to differences in computational platform, and a few other minor setup differences noted in Section 3.
- The HNR2 forecast skill is shown to be similar to HR20, with very few statistically significant differences in track and intensity. Several statistically significant differences favoring HNR2 were noted in storm structure but their magnitude is small compared to the actual errors.
- The worst track and absolute intensity forecasts (outliers) were identified so that forecast improvements for these poorly performing cases can be addressed in the future.
- Model output files have been archived and are available to the community for future studies. Forecast maps and verification graphics, along with this report and additional information are available in the DTC website.

2. Introduction

The DTC performed testing and evaluation for the Hurricane WRF system, known as HWRF (Gopalakrishnan et al. 2010). HWRF was configured as close as possible to the operational HWRF model, employing the same domains, physics, coupling, and initialization procedures as the model used at the NOAA NCEP Central Operations and by the model developers at NCEP EMC. The configuration tested matches the 2011 HWRF Baseline, which is the configuration that served as control for all developments at EMC geared towards the 2011 operational implementation.

The HWRF System has the following components: WPS, prep_hybrid (WRF preprocessor for input of GFS spectral data in native coordinates and binary format), vortex relocation and initialization, GSI 3D-Var, WRF model using a modified NMM dynamic core, POM, features-based ocean initialization, WPP, GFDL vortex tracker, GrADS-based graphics, and NHCVx. HWRF is currently designed for use in the North Atlantic and Eastern North Pacific basins. Atlantic forecasts are in coupled ocean-atmosphere mode, while Pacific forecasts use only the atmospheric model.

3. Goals

The overarching goal of the HWRF 2011 Baseline Test Plan was to establish the skill of the community HWRF code for tropical storm forecasting, designating a new DTC RC (Wolff et al. 2010). This RC, formally referred to as HWRF PS:85.98.98.88.88.2.4 and dubbed HNR2, is now a control against which future advancements in HWRF forecast skill can be measured.

Another goal of this test was to compare the forecasts against those obtained at NOAA EMC using the HR20 configuration of HWRF, in order to assess how similar the forecasts produced with the community code are to forecasts produced with code from the EMC repository. It was not expected that HNR2 and HR20 forecasts would match exactly. Work was conducted to merge the EMC and community code repositories for the WPS, prep_hybrid, vortex relocation, POM, POM initialization, and WRF codes, and previous tests have shown that they now produce identical results. However, minor differences remain between the codes for postprocessing and tracking. Additionally, there are differences in computing platform, in the time-invariant snow albedo dataset, and a bug in the scripting for the ocean was discovered after the HR20 runs were complete and has been fixed in the DTC version of the scripts. While differences between HNR2 and HR20 forecasts are expected to be in the noise level for cold start runs, the cycling nature of HWRF will sometimes cause an amplification of the differences for the later initializations of a given storm. It was expected that while HNR2 and HR20 results would differ for specific storms, the forecast skill would be similar over a sufficiently large sample of cases. Table 1 summarizes the differences between HNR2 and HR20.

Table 1. Differences between HNR2 and HR20

	HNR2	**HR20**
Institution conducting test	DTC	EMC
Computer platform	Linux Cluster *tjet*	NCEP IBMs *CCS* and *vapor*
Source code repository	Community	EMC
Scripts	DTC	EMC
Automation	NOAA GSD Workflow Manager	EMC HWRF History Sequence Manager
I/O format	NetCDF	Binary
WPP	WPP v3.2	NAM Post Modified for HWRF
Tracker	Community repository	EMC operational tracker
Sharpening procedure in ocean initialization for Atlantic domain	Used in ocean spin up Phases 3 and 4 and in coupled model run	Used in ocean spin up Phase 3 only (known bug)
Snow Albedo	Older dataset	Newer dataset

4. Experiment Design

The end-to-end system was composed of WPS, prep_hybrid, vortex relocation and initialization, GSI, ocean initialization, POM, WRF, coupler, WPP, tracker, graphics generation, data archival, and dissemination of results.

a. Codes employed

The software packages utilized were obtained from the community repositories for all codes, except fore prep_hybrid and NHCVx, which are not currently supported to the community. NHCVx will be obtained from a DTC in-house code repository. Prep_hybrid, although not supported to the community, is kept in the hwrf-utilities repository. The revisions for all codes are listed below:

- WRF - https://svn-wrf-model.cgd.ucar.edu, revision 4594
- WPS – revision 573
- WPP – official release v3.2
- GSI – official release v2.5
- Vortex relocation and initialization, prep_hybrid, miscellaneous libraries and tools: https://svn-dtc-hwrf-tne.cgd.ucar.edu, revision 173

- POM and POM initialization – https://svn-dtc-pomtc.cgd.ucar.edu, revision 60
- Coupler - https://svn-dtc-ncep-coupler.cgd.ucar.edu, revision 35
- Tracker - https://svn-dtc-gfdl-vortextracker.cgd.ucar.edu, revision 49
- NHCVx - https://svn-dtc-nhcvx.cgd.ucar.edu, revision 3

The scripts will be obtained from the DTC in-house repository at https://svn-dtc-hwrf-tne.cgd.ucar.edu, revision 83.

b. Domain Configurations

The HWRF domain was configured the same way as used in the NCEP/EMC operational system. The atmospheric model employed a parent and a movable nested grid. The parent grid covered a 75x75° area with 0.18° (approximately 27 km) horizontal grid spacing. There were a total of 216 x 432 grid points in the parent grid. The nest covered a 5.4 x 5.4° area with 0.06° (approximately 9 km) grid spacing. There were a total of 60 x 100 grid points in the nest. Both parent and nest used the WRF-NMM rotated latitude-longitude projection and the E-staggered grid. Indices in the E-staggered grid are such that a square domain has approximately twice as many points in the y-direction than the x-direction. The location of the parent and nest, as well as the pole of the projection, varied from run to run and were dictated by the location of the storm at the time of initialization. Forty-two vertical levels (43 sigma entries) were employed, with a pressure top of 50 hPa.

HWRF was run coupled to the POM ocean model for Atlantic storms and in atmosphere-only mode for East Pacific storms. The POM domain for the Atlantic storms depended on the location of the storm at the initialization time and on the 72-h NHC forecast for the storm location. Those parameters defined whether the East Atlantic or United domain of the POM was used. Both POM domains covered an area from 10.0°N to 47.5°N in latitude, with 225 latitudinal grid points. The East Atlantic POM domain ranged from 60.0° W to 30.0° W longitude with 157 longitudinal grid points, while the United domain ranged from 98.5° W to 50.0° W with 254 longitudinal grid points. Both domains had horizontal grid spacing of approximately 18 km in both the latitudinal and longitudinal directions. The POM used 23 vertical levels and employed the terrain-following sigma coordinate system.

Additional intermediate domains were used for the atmospheric model during the vortex relocation and initialization procedures (see Bao et al. 2010), and during postprocessing (see item 3.g below).

Figure 1. Sample domains for the atmospheric (yellow lines outline the stationary outer domain and the moving nest) and oceanic (blue line outlines the United OM domain) components of HWRF.

c. Forecast cases and initial and boundary Conditions

Forecasts were initialized every 6 hours for 53 storms and were run out to 126 hours. When no Best Track information was available, verification was not performed. Additionally, forecasts for observed storms classified as a wave or low by the NHC were excluded. Therefore, the number of verified cases is slightly smaller than the total number of cases.

Initial Conditions were be based on pre13d GFS analysis. Pre13d GFS refers to the retrospective runs of the T574 GFS which was implemented operationally on July 28, 2010. The IC and BC for the atmosphere were obtained from the binary spectral GFS files in native vertical coordinates using prep_hybrid. The IC for the surface fields were obtained from the 1x1° GFS files in GRIB format using WPS. HWRF used a vortex relocation procedure as described in Bao et al. 2010 and Gopalakrishnan et al. 2010. In the presence of a 6-h forecast from a HWRF run initialized 6-h before the initialization time for a given cycle, the vortex relocation procedure removed the vortex from the GFS analysis and substituted it with the vortex from the previous HWRF forecast, after correcting it using the observed location and intensity. When a previous HWRF forecast was not present, the GFS vortex was removed and substituted by a synthetic vortex derived from a procedure that involves theoretical considerations and HWRF climatology. This procedure is referred to as *cold start*.

Typically a cold-start initialization was employed for the first NHC Storm Message (INVEST) of each storm, and the HWRF vortex was cycled for all subsequent initialization of each storm. When the NHC storm message was missing during a storm, there was an interruption in the cycling, and a new cold start was done.

For storms classified as *deep* by the NHC at the time of model initialization, the HWRF initialization was updated using GSI. The data supplied to GSI consisted of conventional restricted prepbufr observations, satellite observations from NOAA, metop-a, AQUA, GOES, and AMSU A and B satellites. For any given analysis, only a subset of the observations were employed because of quality and availability of the datasets. No data was assimilated in the inner core of the storm, that is, the GSI modifications to the HWRF initialization were only applied to the storm environment (outside 150 km radius from the storm center).

The inventory of cases used in this test is listed in Appendix A. All the cases listed in the test plan were run. Three cases could not be completed:

- Bill 03L of 2009 initialized at 06 UTC on 08/23. Executable copygb crashed for lead times 84 h and beyond due to a well known deficiency in this program for cases in which the center of the nest is very distant from the center of its projection. After Bill went extratropical and quickly moved to the North and East, copygb failed.
- Gaston 09L of 2010 initialized at 18 UTC on 09/07. This case could not be initialized in cycled mode because storm in the previous case (12 UTC on 09/07) was too weak and the tracker only ran for the first 6 h. A cold start was not attempted.
- Eleven 11E of 2010 initialized at 12 UTC on 09/04. The vortex initialization procedure failed, probably because the storm was over the steep topography of central America.

This failures will be investigated with the intention of making the modeling system more robust. In addition, archives were lost for five cases of Ike 09L of 2008 (00 UTC on 09/13 through 00 UTC on 09/14).
Forecast verification was not performed for all cases run. For some cases Best track information was not available. Additionally, verification was not performed if the storm was classified as a low or wave.

d. Physics Suite

The physics suite configuration (Gopalakrishnan et al. 2010) is described in Table 2. The convective parameterization was applied in both the parent and nest domains with momentum mixing (coefficient 1.0) activated in both.

Table 2. Physics Suite for HNR2 test.

Microphysics	Ferrier for the tropics (85)
Radiation SW/LW	GFDL/GFDL (98/98)
Surface Layer	GFDL (88)
Land Surface Model	GFDL slab model (88)
Planetary Boundary Layer	GFS (3)
Convection	Simplified Arakawa-Schubert (4)

e. Other aspects of code configuration

The HWRF system was compiled with the environmental variables WRF_NMM_CORE, WRF_NMM_NEST and HWRF set to 1 in order for the executables to contain the HWRF-specific instructions.

As in the operational configuration, a time step of 54 s was used for the parent grid, while a time step of 18 s was used in the nest. Calls to the turbulence, cumulus parameterization and microphysics were done every 4.5 minutes for the parent domain and 54 s on the nest. Calls to the radiation were done every 54 minutes on the parent grid and 9 minutes on the nest. Coupling to the ocean model and nest motion were restricted to a 9-minute interval.

The gravity wave drag parameterization was applied in the parent-domain only, and the advection used the Lagrangian scheme.

f. Post-processing and Vortex Tracking

The wrfpost program within WPP was used in the parent and nest domains to destagger the forecasts, to generate derived meteorological variables, including MSLP, and to vertically interpolate the fields to isobaric levels. The post-processed fields included two- and three-dimensional fields on constant pressure levels and at shelter level, all of which are required by the plotting and verification programs.

Using the copygb program contained in WPP, the post-processed parent and nest domains were horizontally interpolated to a latitude-longitude grid with similar domain size to the parent domain and grid spacing similar to the native nested domain. Those two grids with same domain and grid spacings were then combined in order to create a high-resolution grid covering an area similar to the parent domain. Additionally, the post-processed forecast from the nest domain was horizontally interpolated to a high-resolution standard latitude-longitude grid with similar domain to the nest in order to generate graphics.

Tracking was done on the combined domain. For purposes of verification and graphics generation, the input was six hourly postprocessed files. Tracking for the purposes of cycling the HWRF vortex used three-hourly postprocessed files.

g. Model Verification

All verification graphics can be seen at http://verif.rap.ucar.edu/eval/hwrf_hnr2_hr20/verify/. The characteristics of the forecast storm (location, intensity, structure) as contained in the HNR2 and HR20 ATCF files produced by the tracker were compared against the Best Track using the NHCVx. The HNR2 ATCF files were produced by the DTC as part of this test, while the HR20 ATCF files were supplied by EMC. The NHCVx was run separately for each case, at 6-hourly forecast lead times, out to 120 h, in order to generate a distribution of errors.

Verification was performed for any geographical location for which Best Track was available, including over land. No verification was performed when the observed storm was classified as a low or wave.

A R-statistical language script was run separately on an homogenous sample of the HNR2 and HR20 datasets to aggregate the errors and to create summary metrics including the mean and median of track error, intensity error, absolute intensity error, and radii of 34, 50, and 64 kt wind in all four quadrants. All metrics are accompanied of 95% confidence intervals to describe the uncertainty in the results due to sampling limitations. The largest outliers (worst forecasts) were identified.

It was originally planned that along- and cross-track errors would be included in this evaluation. However, an error was found in the program after the errors were computed and therefore they are not included in this report.

For the purposes of comparing the HNR2 and HR20 forecasts, pairwise differences (HNR2-HR20) of track error, intensity error and absolute intensity error were computed and aggregated with a R-statistical language script. Ninety-five percent confidence intervals on the median were computed to determine if there is a statistically significant difference between the two configurations.

For the intercomparison between HNR2 and HR20, a homogeneous sample was used. On the other hand, for the analysis of the individual errors of the HNR2 and HR20 configurations, this constraint was not enforced. However, the heterogeneity of the sample is minimal because the HNR2 and HR20 sets cover the same cases.

Verification results are described in Section 7.

h. Graphics

All forecast graphics can be seen at http://www.dtcenter.org/HurrWRF/graphics/HNR2-HR20.

Graphics were generated using GrADS scripts originally developed at EMC. Graphics included line plots of track, maximum winds and mean sea level pressure.

Additionally, the following 4 graphics were produced for six-hourly lead times

- 850-hPa streamlines and isotachs on the combined domain
- 850-hPa streamlines and isotachs on the nest
- MSLP and 10m winds on the nest
- Zonal cross sections of zonal and meridional wind on the nest
- Meridional cross section of zonal wind on the nest

i. Archives

Input and output data files from several stages of the end-to-end system have been archived in the NOAA ESRL/GSD MSS.

The input GFS spectral data in binary format data, along with the observations used in GSI, can be found at /arch3/jet/projects/dtc-hurr/datasets/*yyyymmdd*, where *yyyymmdd* is the year, month and day of the forecast initialization.

The TCVitals, the A- and B-decks (containing HR20 tracks), the files for ocean initialization (Loop current and warm and cold core rings) along with all the fix (static) files can be found in /arch3/jet/projects/dtc-hurr/HWRF_HNR2_run_archive/dataset.tar.bz2.

The output can be found at /arch3/jet/projects/dtc-hurr/HWRF_HNR2_run_archive/*SID_yyyymmddhh*.tar.bz2, where *SID* is the Storm Identification, expressed as 2 digits plus one letter (L for Atlantic and E for East Pacific). Appendix B lists all the files that are archived for each case.

File /arch3/jet/projects/dtc-hurr/HWRF_HNR2_run_archive/tracks_HNR2.tar contains all the track files for HNR2.

A file with the output from the NHC Vx for all HNR2 and HR20 cases can be found at /arch3/jet/projects/dtc-hurr/HWRF_HNR2_run_archive/HWRF_HNR2_NHC_files.tar.bz2.

The scripts used in the postprocessing of the NHC Vx data, along with all the logs and images produced, can be found at /arch3/jet/projects/dtc-hurr/HWRF_HNR2_run_archive/HNR2_HR20_Rscript_and_output.tar.bz2

All source codes and executables are in /arch3/jet/projects/dtc-hurr/HWRF_HNR2_run_archive/TNE_source_files.tar.bz2.

5. Computer resources

- Processing resources

  All forecasts were computed on the HFIP Linux cluster *tjet* located at NOAA GSD. For the coupled run, 91 processors were used for the atmospheric model, 1 for the coupler, and 1 for POM. GSI was run in 24 processors. All other programs were run in a single processor.

- Storage resources

All archival are on the NOAA GSD MSS.

- Web resources

  Model forecast and verification graphics can be accessed through a web interface available on the DTC website.

6. Deliverables

The NOAA GSD MSS was used to archive the files input and output by the forecast system. Appendix B lists the output files that will be archived. Additionally, all code compilation logs, input files and fixed files used in the runs have been archived. These files are available to the community for further studies.

The DTC website is being used to display the forecast and objective verification graphics.

Finally, this report summarizes the results and conclusions from this test.

7. Results

For brevity, this reports gives a summary of the most important results. A comprehensive set of verification figures is available at http://verif.rap.ucar.edu/eval/hwrf_hnr2_hr20/verify/.

a. North Atlantic Basin

The mean of the track errors for HNR2 and HR20 indicates that the errors grow in time from near zero at the initialization time to 280 nm at the five-day lead time (Fig. 2). There is only one statistically significant (SS) difference in the median of track errors between HNR2 and HR20, occurring at the 78-h lead time (Fig. 3). This difference favors the HNR2 configuration

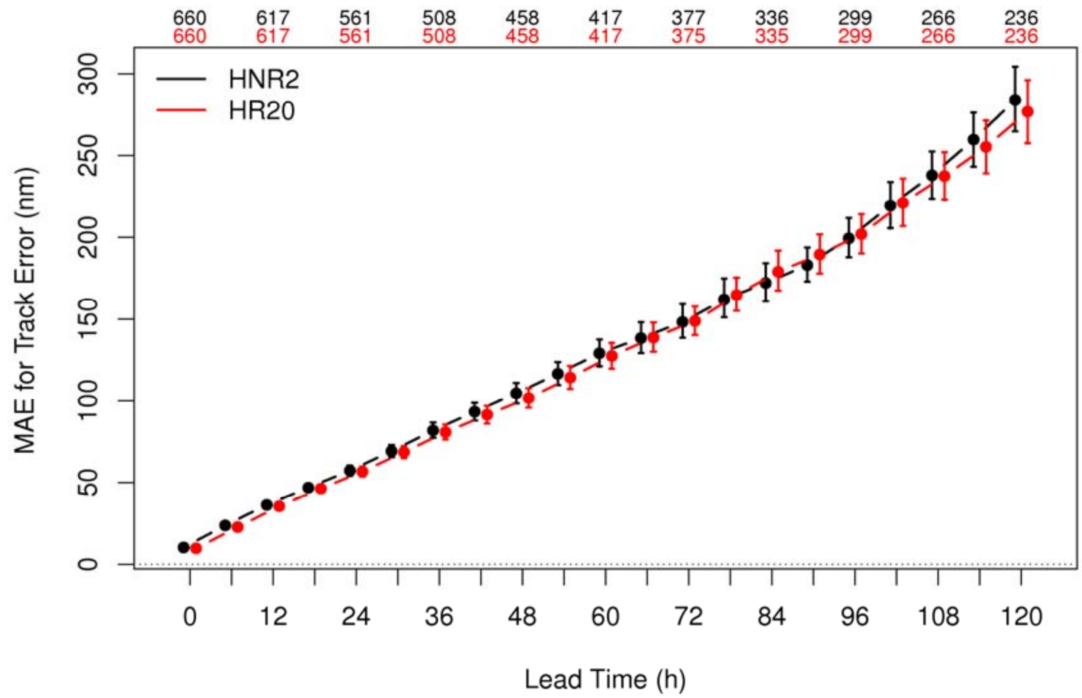

Figure 2. Mean track error (nm) for HNR2 (black) and HR20 (red) as a function of forecast lead time for all cases in the Atlantic basin. The 95% confidence intervals are also displayed. The sample size is listed above the graphic.

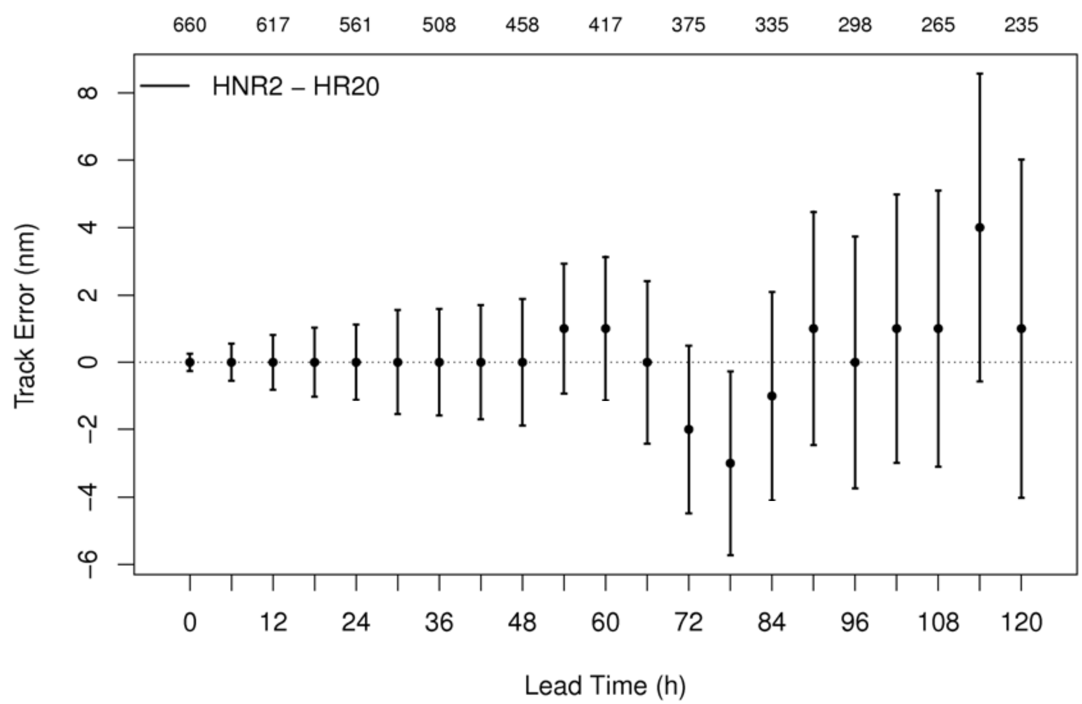

Figure 3 . Median of the pairwise difference in track error (nm) between HNR2 and HR20 as a function of forecast lead time for all cases in the Atlantic basin. The 95% confidence intervals are also displayed. Positive (negative) values indicate HR20 (HNR2) superior performance. The sample size is listed above the graphic.

The mean of the absolute intensity errors for HNR2 and HR20 displays a sharp growth in the first 6 h of forecast, from near zero to 10 kt. From 6 to 48 h, the error grows more slowly to about 16 kt, and remains relatively unchanged after that (Fig. 4). The mean of the intensity errors for HNR2 and HR20 shows a SS negative bias from the initialization up until the 78–h lead time (Fig. 5). The bias is small (approximately -1 kt) at the initial time, but grows sharply to -4 kt after 6 h, suggesting an imbalance in the initialization. The worst bias are seen after 1 day of forecasting, after which the bias diminishes and becomes near zero by the end of the forecast period. There are no SS differences in the median of absolute intensity errors between HNR2 and HR20 (Fig. 6).

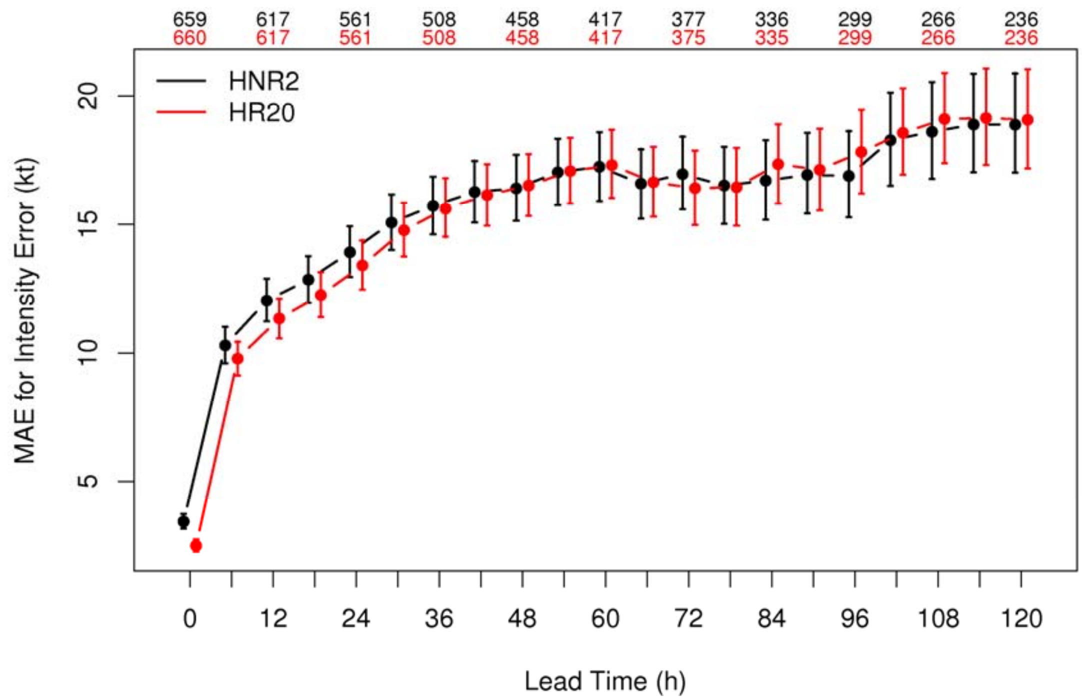

Figure 4. Same as Fig. 2, except for absolute wind error (kt).

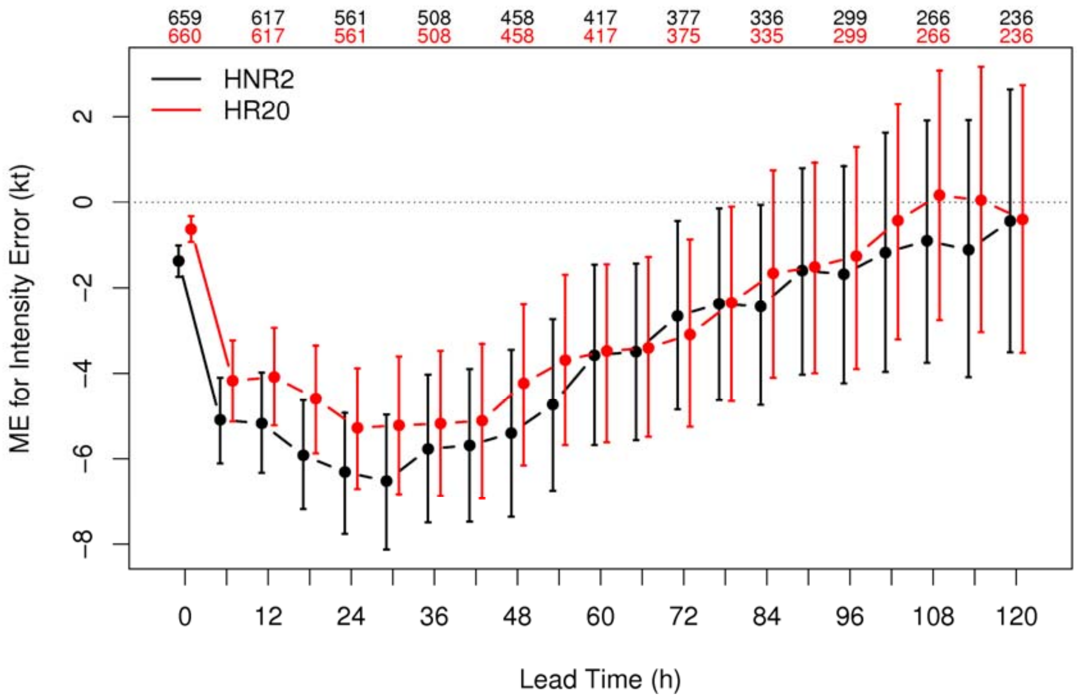

Figure 5. Same as Fig. 2, except for intensity error (kt).

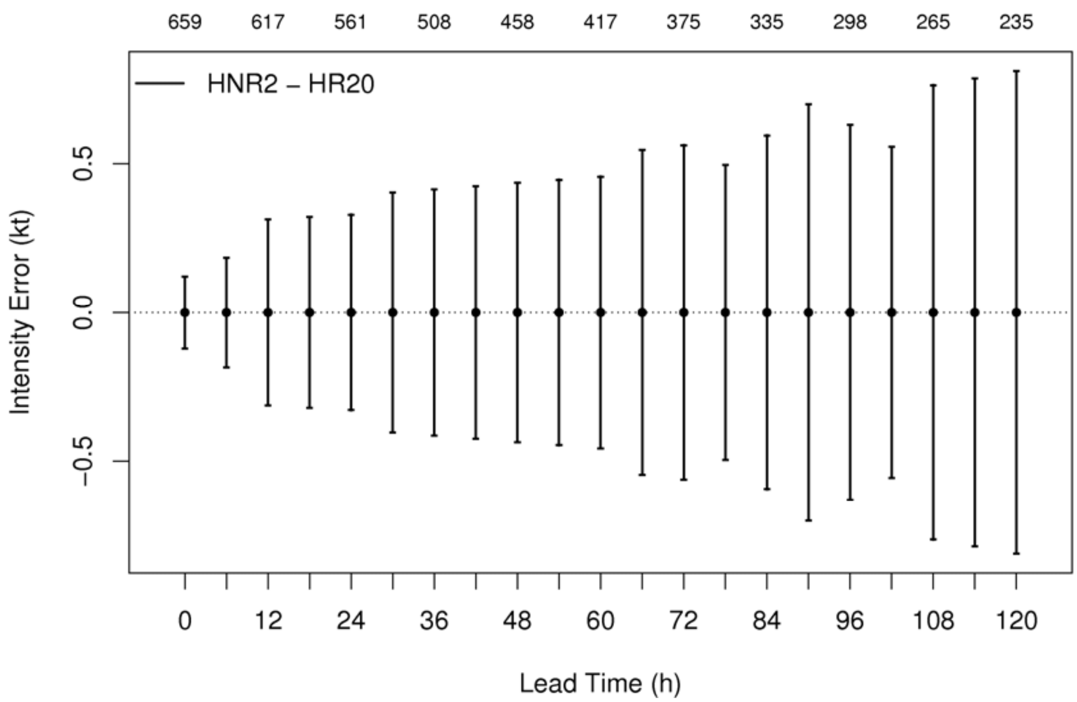

Figure 6. Same as Fig. 3, except for absolute intensity error (kt).

The mean absolute errors for the 34 kt radii are in order of 45 nm, while they are on the order of 30 kt for the other radii (not shown). In general the errors tend to be higher at the beginning and at the end of the forecast period, pointing to deficiencies in the initialization of storm structure. In a couple of cases (34 kt radii for the SE and SE quadrants), the error grows sharply in the first 6 h, then decays, and finally increases towards the end of the forecast, suggesting that the model goes through an adjustment period in the first six hours which is not in alignment with the observed storm structure. The mean error is positive for all quadrants and wind thresholds, indicating that the forecast storm is always larger than the observed one.

The lead times in which SS differences between HNR2 and HR20 in track, intensity and storm structure occur are listed in Table 3. In all cases, the SS differences have small magnitude compared to the errors of the individual models.

While the 34 kt wind threshold in the SW quadrant of the storm does not exhibit any SS difference between HNR2 and HR20, differences are noted in the other quadrants in 33 instances (Table 3) Note that there are a total of 252 instances (21 lead times, 4 quadrants, 3 radii) that can exhibit differences. All differences favor the HNR2 configuration, except for the 50 and 64 kt radii in the NE quadrant at the initialization time, which favor HR20.

Table 3. Statistical significances in the mean of various errors as a function of forecast lead time for all cases in the Atlantic basin. The lines correspond to track error (Tk), intensity (Wd), radii of the 34, 50, and 64 kt wind threshold in the NE, SW, SE, and SW quadrants of the storm (34NE, 50NE, 64NE, 34NW, 50NW, 64NW, 34SE, 50SE, 64SE, 34SW, 50SW and 64SW, respectively). Statistically significant differences that favor the HNR2 or HR20 configurations are indicated as "N" or "R", respectively. A dash indicates that no SS difference exists.

	0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Tk	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-
Wd	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NE34	-	-	-	-	-	N	-	-	-	N	-	-	N	-	N	N	N	N	N	-	-
NE50	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	N
NE64	R	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NW34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-
NW50	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-
NW64	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SE34	-	N	-	-	-	-	-	N	N	N	N	-	-	-	-	-	N	-	N	-	-
SE50	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	N	-	N
SE64	-	-	-	-	N	N	-	-	-	-	-	-	-	-	-	-	-	N	-	N	-
SW34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 7 displays the outliers in track forecasting for the Atlantic Basin for the HNR2 configuration. The worst forecasts in the first two days of forecast are from Nicole, while Thomas is responsible for several outliers on days 3 and 4. Finally, the worst 5-day forecasts are from Fay.

Even though the largest track error outliers happened for Nicole, Thomas and Fay, those are not necessarily the worst storms on average. When the error is computed individually for each storm, the largest mean track errors are for Erika in which the error reached 300 nm in 48 h, Ida with 550 nm in 5 days, Matthew with 300 nm in 60 h, and Nicole with 400 nm in 42 h.

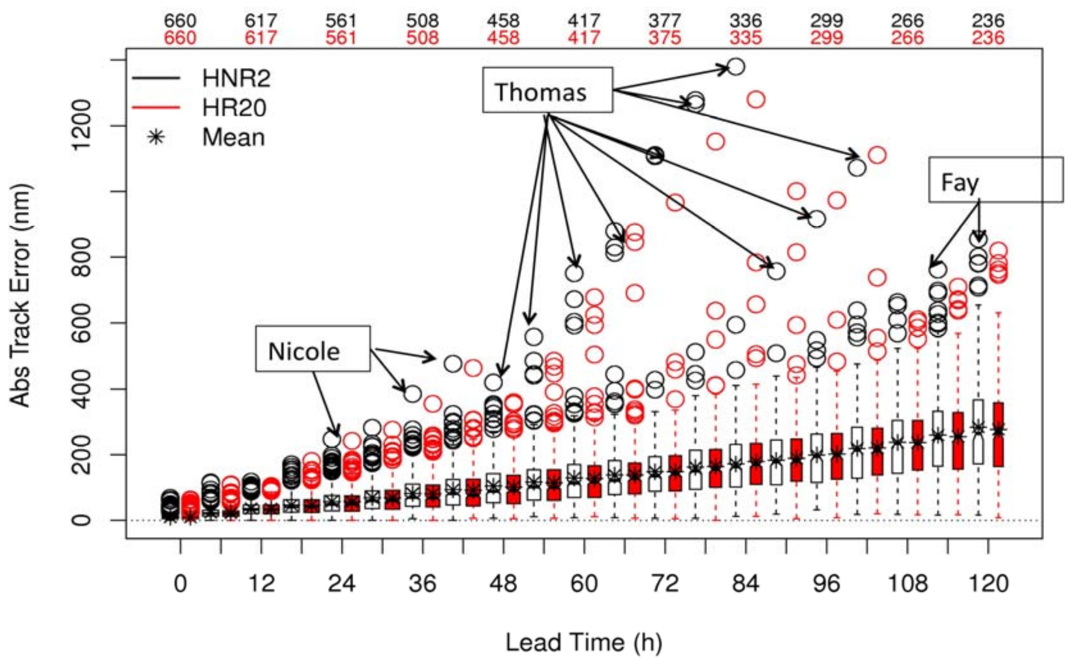

Figure 7. Boxplots of mean track errors for the HNR2 (black) and HR20 (red) configurations. The median is the waist of the plot and the 95% confidence intervals are the notches. The bottom and top of the bocplots denote the 25 and 75% percentiles, respectively. Outliers are represented as circles. A star represents the mean.

The outliers for absolute intensity can be seen in Fig. 8. A large number of outliers is seen in general for the intensity forecasting, indicating that there is a strong variability in the forecast skill from run to run. The worst forecasts in the first two days are from Igor, while large outliers for Gustav, Ike, and Richard are seen later in the forecast period. The intensity MAE did not present a large variability between storms, that is, it was not possible to identify individual storms that had much worst absolute intensity forecast than the other storms.

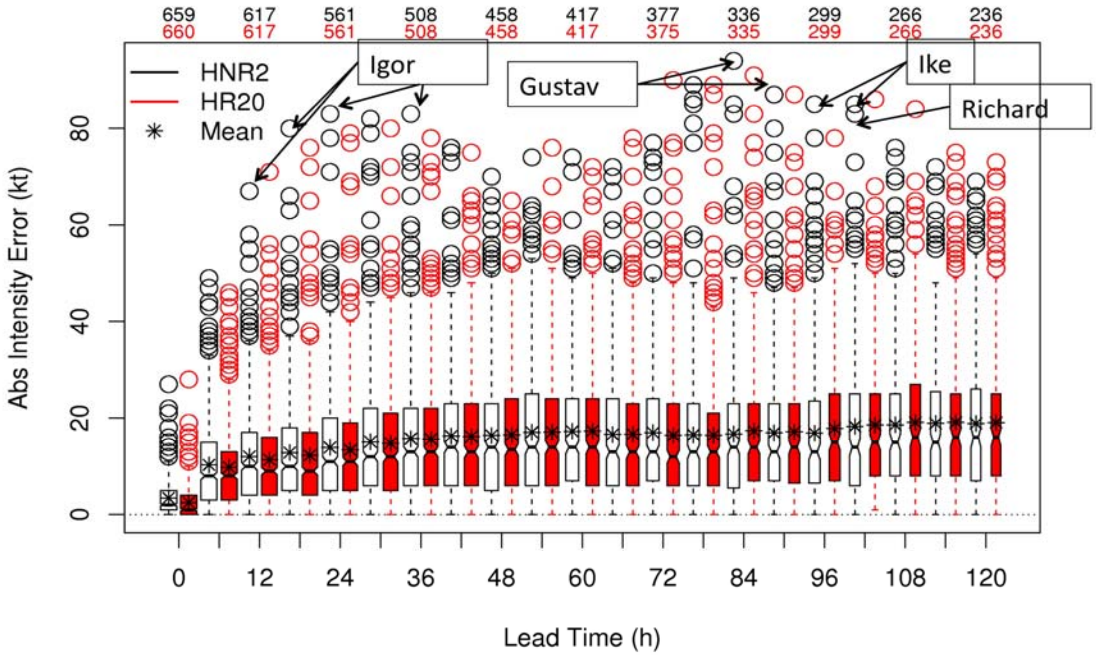

Figure 8. Same as Fig, 7 but for absolute intensity.

b. Eastern North Pacific Basin

Similarly to the results in the Atlantic basin, the track errors in the Eastern North Pacific basin increase linearly with time during the forecast period, from near zero at the initial time to approximately 200 nm for the 5-day forecast (Fig. 9). Therefore, the Pacific errors are smaller than the Atlantic ones, indicating that the forecast challenge is easier in the Pacific. Statistically significant differences in track error between HNR2 and HR20 appear at 48- and 60-h lead times, both favoring HR20, but in both cases with magnitude smaller than 3 nm (Fig. 10).

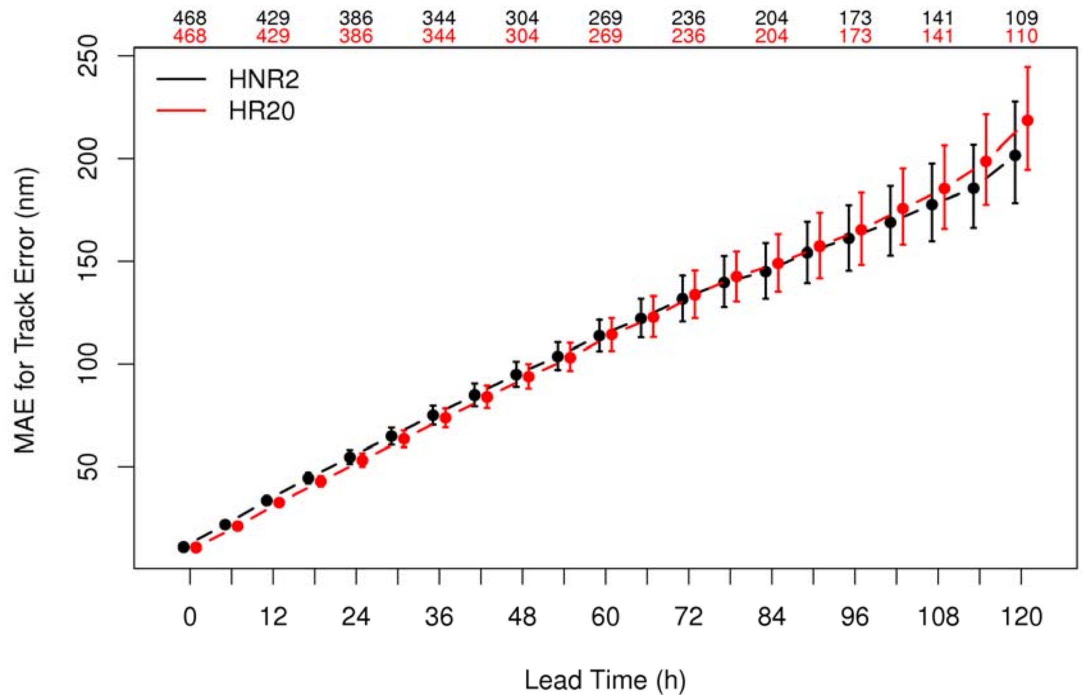

Figure 9. Mean track error (nm) for HNR2 (black) and HR20 (red) as a function of forecast lead time for all cases in the North Pacific basin. The 95% confidence intervals are also displayed. The sample size is listed above the graphic.

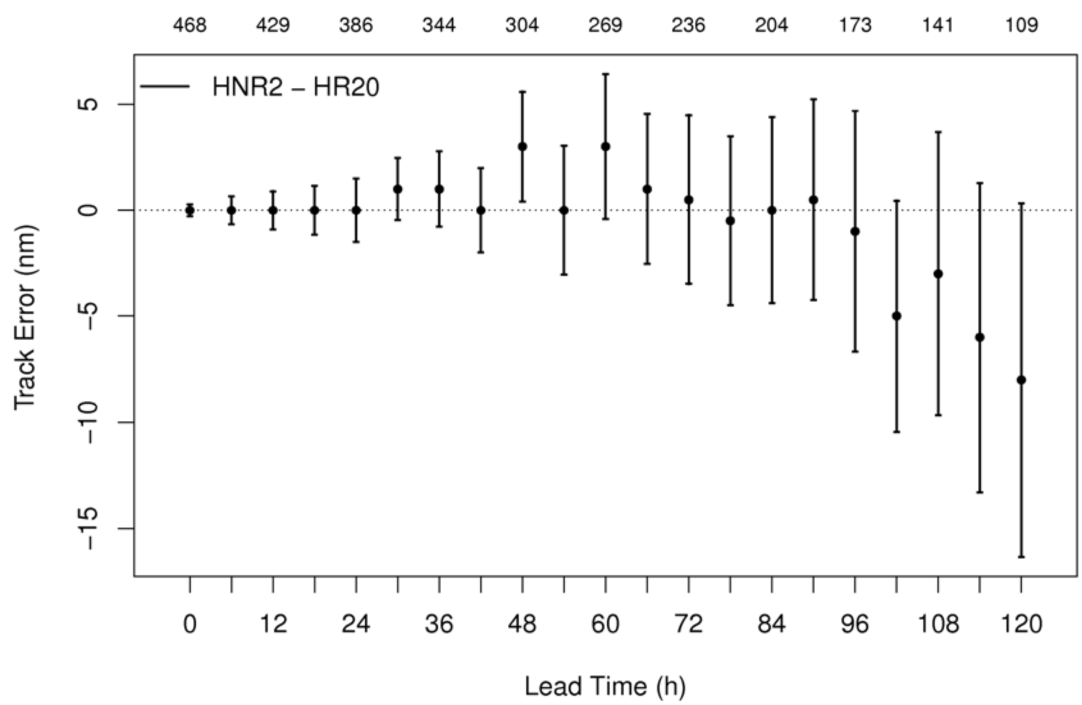

Figure 10. Median of the pairwise difference in track error (nm) between HNR2 and HR20 as a function of forecast lead time for all cases in the North Pacific basin. The 95% confidence intervals are also displayed. Positive (negative) values indicate HR20 (HNR2) superior performance. The sample size is listed above the graphic.

The mean absolute intensity error is also similar to the Atlantic basin, with errors increasing very rapidly in the first 6 h, then grows slowly until the third day of the forecast, becoming stable after that (Fig. 11). The mean intensity errors (Fig. 12) indicate that the error is near zero at the initial time, and grows sharply in the first six hour of forecast to develop a negative bias which remains throughout the entire forecast period. Statistically significant differences in storm intensity (Fig. 13) only occur at the 48- and 72-h forecasts, favoring HR20 and HNR2 respectively. The magnitude of the difference does not exceed 1 kt.

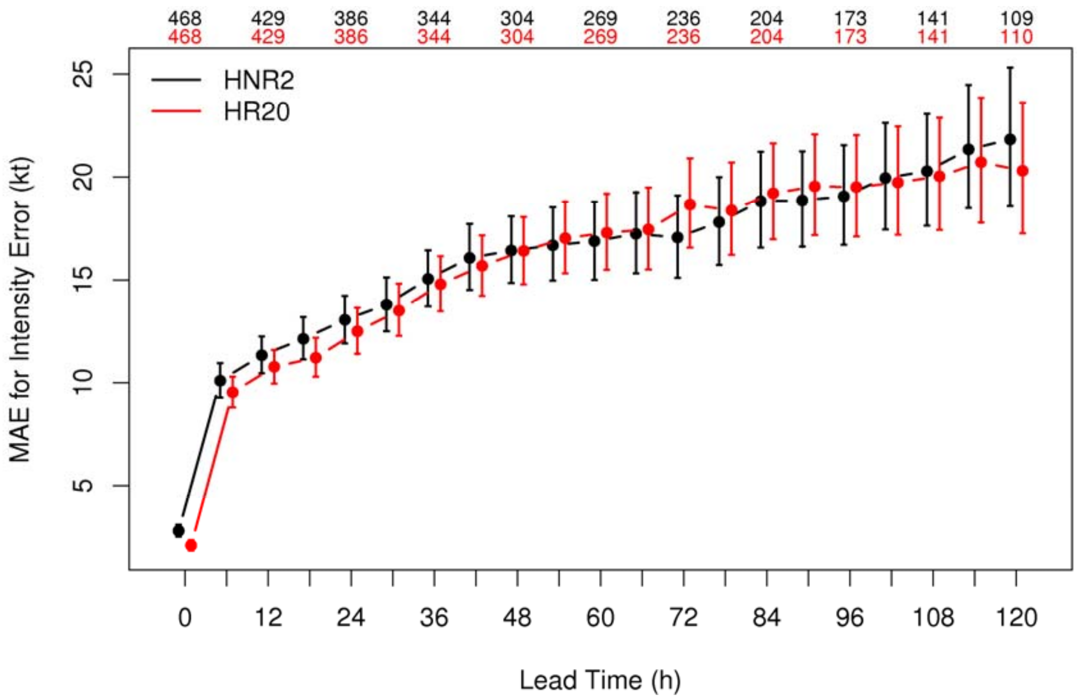

Figure 11. Same as Fig. 6 but for absolute intensity error (kt).

The forecast storm size, as defined by the radii of the 34-, 50-, and 64 kt wind thresholds, is consistently larger than the observed one. For the 50- and 64 kt thresholds, the best forecast is obtained at the 24-h lead time, that is, the errors decrease in the first day of forecast pointing to an initialization issue. On the other hand, for the 34-kt threshold in most quadrants, the errors increase in time in the first 6-h, and decrease thereafter in the first day of forecast, indicating that the initialization problem gets exacerbated in the first few hours of the forecast.

All SS differences between HNR2 and HR20 are summarized in Table 4. Numerous SS differences occur between HNR2 and HR20 in the radii of the various wind thresholds (Table 4), indicating that there might be a systematic difference in storm structure between the two configurations. However, the differences are of small magnitude when compared to the actual errors. All differences except one favor the HNR2 configuration, indicating that HNR2 produces a more compact storm than HR20.

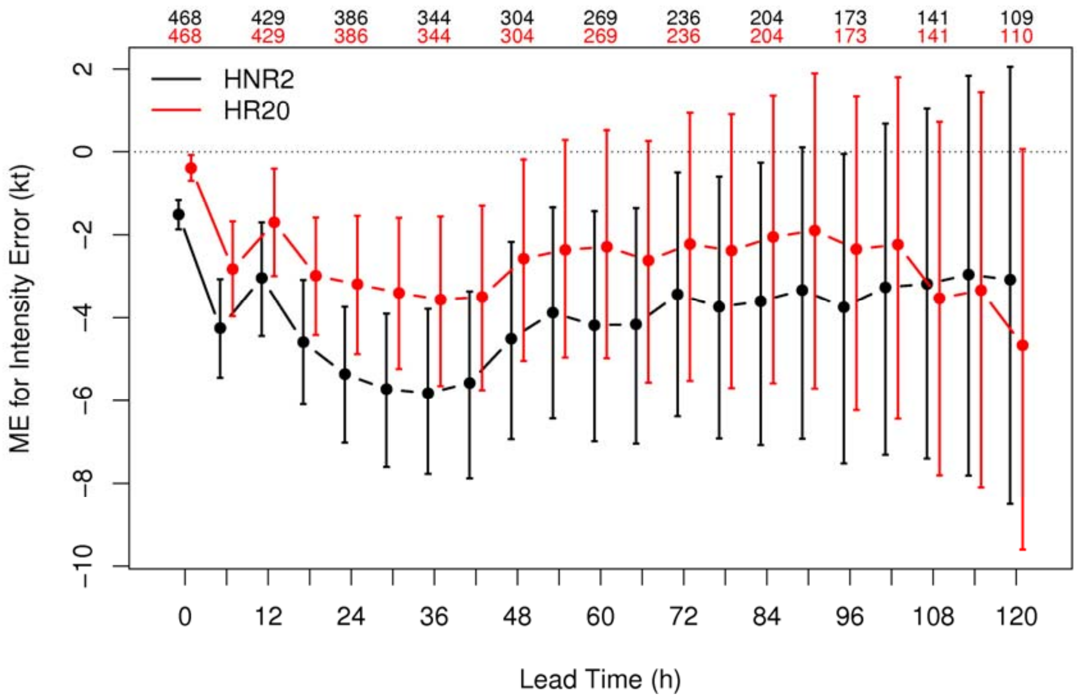

Figure 12. Same as Fig. 6, but for intensity errors (kt).

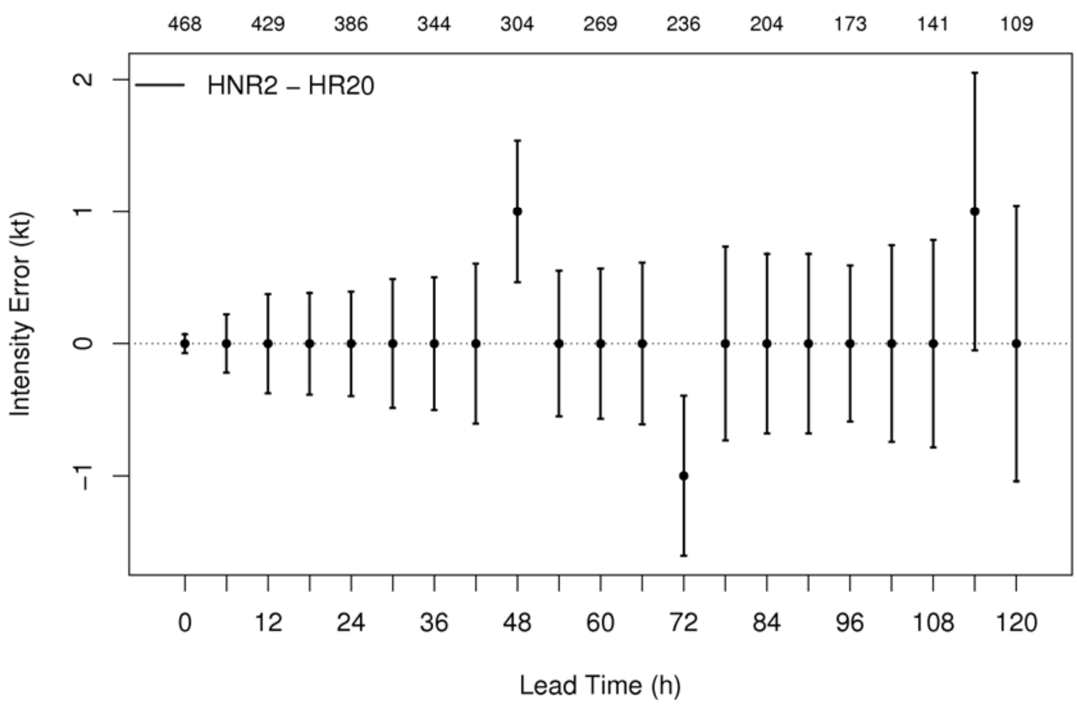

Figure 13. Same as Fig 7 but for absolute intensity errors.

Figure 14 displays the outliers in track forecasting for the Eastern North Pacific Basin for the HNR2 configuration. The worst forecasts in the first two days of forecast are from Olaf, while Darby is responsible for several outliers on days 3 through 5.

When the errors are aggregated for individual storms, the largest mean track errors are for Olaf in which the error surpassed 250 nm in 48 h, Rick and Darby with more than 450 nm in 5 days, and Estelle, which reached 300 nm in 102 h.

The outliers for absolute intensity can be seen in Fig. 15 The worst forecasts in the first two days are from Rick, Celia, and Guillermo, while large outliers for Celia dominate later in the forecast period. When the MAE is computed separately for each storm, the worst intensity forecasts are for Rick, in which the error reaches 40 kt in 48 h, and for Celia, that surpasses 30 kt in the first day of day forecast (not shown).

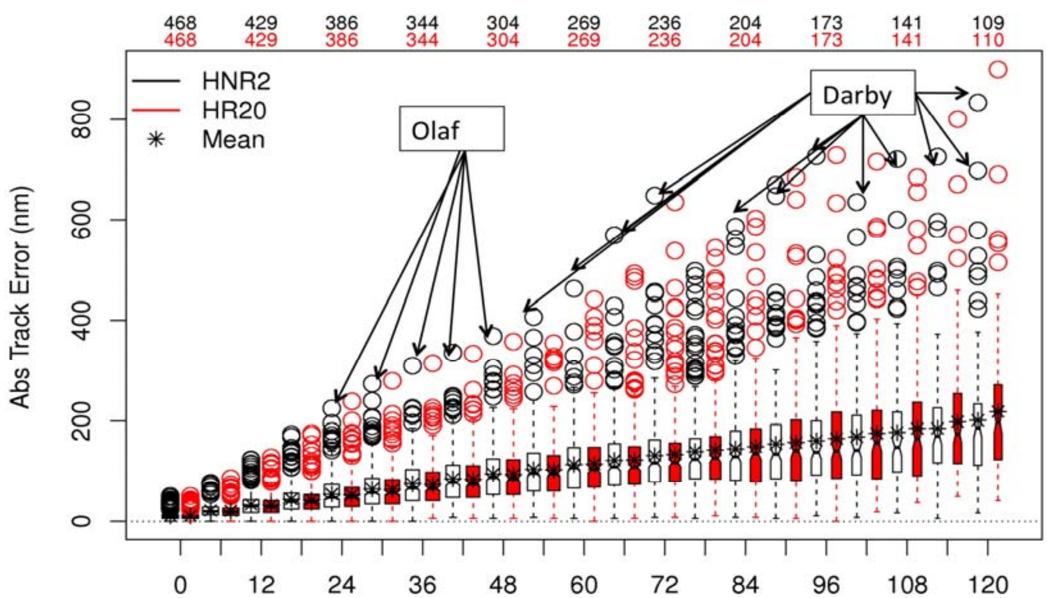

Figure 14. Boxplots of mean track errors for the HNR2 (black) and HR20 (red) configurations for the Eastern North Pacific Basin. The median is the waist of the plot and the 95% confidence intervals are the notches. The bottom and top of the bocplots denote the 25 and 75% percentiles, respectively. Outliers are represented as circles. A star represents the mean.

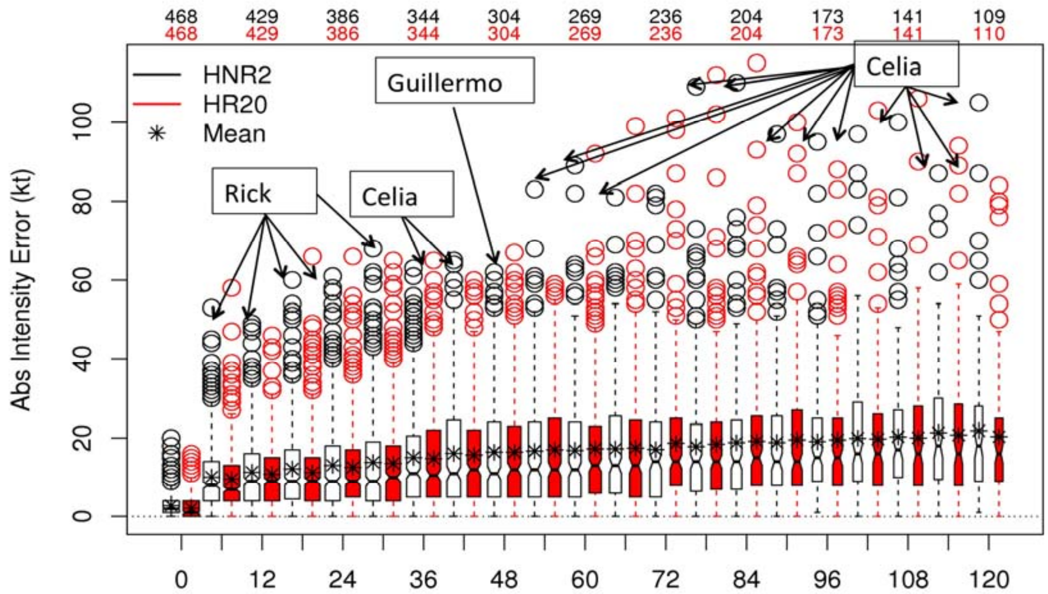

Figure 15. Same as Fig. 14 but for the absolute intensity errors.

Table 4. Same as Table 3, except for the Eastern North Pacific basin.

	0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Tk	-	-	-	-	-	-	-	-	R	-	-	-	-	-	-	-	-	-	-	-	-
Wd	-	-	-	-	-	-	-	-	R	-	-	-	N	-	-	-	-	-	-	-	-
NE34	-	N	N	N	N	N	-	N	-	N	N	-	N	N	N	-	-	-	-	-	-
NE50	-	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	-	-
NE64	-	N	-	N	N	N	N	N	N	N	N	-	N	N	N	N	N	N	-	N	-
NW34	-	-	-	N	-	N	-	N	-	-	-	N	-	N	N	-	-	-	-	-	-
NW50	-	-	-	-	-	-	-	-	-	N	-	-	N	-	-	-	-	-	-	-	-
NW64	R	-	-	-	-	-	-	-	N	-	-	-	N	-	N	-	N	N	-	-	-
SE34	-	N	N	N	N	N	N	N	N	N	N	N	N	-	-	-	-	-	-	N	-
SE50	N	N	N	N	N	N	N	N	N	N	-	N	-	N	N	N	N	N	N	-	N
SE64	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	-	-
SW34	-	N	N	N	N	N	N	N	N	N	N	N	N	-	-	-	-	-	-	-	-
SW50	-	-	-	N	-	N	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-
SW64	-	-	-	N	N	-	-	-	-	-	N	N	N	-	-	-	-	-	-	-	-

8. Discussion and Conclusions

This was the first extensive test of HWRF configurations conducted by the DTC, and it demonstrated the DTC's ability to efficiently conduct extensive HWRF experiments in an environment that is functionally equivalent to EMC's. The runs were very robust, however three crashes occurred. For storm Bill 03L of 2009, initialized at 08/23 at 06 UTC, copygb crashed at the 84-h and later lead times. At this time, Bill had undergone extratropical transition and had quickly moved north and east, placing the center of the nest far from the center of the parent grid projection. This is a known challenge for the copygb program, which we recommend be amended to better handle this situation. The second crash occurred for storm 11E, initialized at 09/04 at 12 UTC. The center of the storm at the time of initialization was over land and high topography, which caused the vortex initialization procedure to produce unrealistic results. It would be worth it to revise the vortex initialization procedure to better handle the vortex relocation and correction procedure over complex topography. Finally, the Gaston (09L of 2010) forecasts initialized at 09/07 at 18 UTC and 09/08 at 00 UTC did not run because the storm in the forecast initialized at 09/07 was too weak to be tracked and produce results to be used in the subsequent cycled start.

The forecast results indicate that the HNR2 configuration run by DTC produced similar results to the HR20 configuration run by EMC, with a minor number of SS differences in track and intensity. A larger number of SS differences between HNR2 and HR20 was found in the radii of the 34-, 50-, and 64-kt wind thresholds, mostly favoring the HNR2 configuration, indicating that there may be a systematic difference between the two configurations. However, given the small magnitude of the differences, exploring this issue was not given a high priority.

The main purpose of the test, the benchmarking of the HWRF community code, was achieved and the HWRF PS:85.98.98.88.88.2.4 RC, nicknamed HNR2, was created and results were published. This RC will serve as a baseline against which future RCs will be compared to assess the improvement of the HWRF model.

9. References

Bao, S., R. Yablonsky, D. Stark, and L. Bernardet, 2010. HWRF Users' Guide. Developmental Testbed Center, 88pp.

Gopalakrishnan, S., Q. Liu, T. Marchok, D. Sheinin, N. Surgi, R. Tuleya, R. Yablonsky, and X. Zhang, 2010: Hurricane Weather and Research and Forecasting (HWRF) Model scientific documentation. L. Bernardet, Ed., 75 pp.

Wolff, J., L. Nance, L. Bernardet and B. Brown, 2010. Implementation of WRF Reference Configurations. WRF Workshop, Boulder, CO.

Acknowledgements
The Developmental Testbed Center is funded by the National Oceanic and Atmospheric Administration, Air Force Weather Agency and the National Center for Atmospheric Research. This work was supported by the NOAA Hurricane Forecast Improvement Project

Appendix A: Inventory

Table 3. Inventory for HNR2 Test. Columns on the table refer to the storm name, storm number, beginning and ending case (month, day and time UTC in format mmddhh), number of cases for which the NHCVx was run, and number of cases for which the NHC Vx contains valid data. Typically the first case of a storm was initialized as a cold start and subsequent cases are cycled. When the NHC storm message was missing during a storm, there was an interruption in the cycling, and a new cold start was done. This is indicated on the table by using multiple lines for a single storm.

2008 Atlantic		Begin date	End date	# nhcvx files	# filled nhcvx files	Notes
Fay	06L	081518	082400	34	34	
Gustav	07L	082512	090118	30	30	
Hanna	08L	082806	090700	40	40	
Ike	09L	090112	091218	46	46	Lost vx output for 091300-091400
Total				150	150	
2008 Pacific						
Elida	06E	071206 071518 072212	071506 071918 072300	13 17 3	13 15 0	 LO 071912, 071918 No BT 072212, 072218, 072300
Fausto	07E	071606 071618	071606 072218	26	26	
Genevieve	08E	072112	072712	25	24	LO for 072712
Marie	14E	100106	100618	23	23	
Norbert	15E	100400	101212	35	35	
Total				142	136	
Total 2008				292	286	
2009 Atlantic						
Bill	03L	081518	082318	32	32	Copygb crashed 082306, hours 84 and later
Claudette	04L	081606	081706	5	5	
Danny	05L	082612	082900	11	11	
Erika	06L	090118	090412	12	9	LO 090400 and 090406, No BT 090412
Fred	07L	090718 091518 091812 092000 092012	091218 091600 091912 092000 092018	20 2 5 1 2	20 0 0 0 0	LO 091218 All LO All LO No BT No BT
Henri	10L	100618	100812	8	8	
Ida	11L	110418	111006	23	23	

Total				121	108	
2009 Pacific						
Felicia	08E	080400 081212	081118 081212	32 1	31 0	LO 081118. No BT
Guillermo	10E	081212	082000	31	29	LO 081918, 082000.
Hilda	11E	082212 082906	082812 082912	25 2	25 0	 LO 082906, 082912
Ignacio	12E	082418	082712	12	11	LO 082712
Jimena	13E	082900	090418	28	28	
Linda	15E	090706	091200	20	19	LO 091200
Olaf	18E	100112	100400	11	9	LO 100318, 100400
Rick	20E	101518	102118	25	25	
Total				187	177	
Total 2009				308	285	
2010 Atlantic						
Alex	01L	062600	070106	22	22	
Bonnie	03L	072212	072418	10	10	
Collin	04L	080218	080812	24	17	LO 080400-080512
Danielle	06L	082118	083018	37	37	
Earl	07L	082512	090412	41	41	
Fiona	08L	083100	090318	16	16	
Gaston	09L	090112 090312	090300 090712	7 17	5 3	LO 090218, 090300 LO 090312, 090318, 090400, 090500-090800 Could not run 090718 and 090800 cycled because storm was too weak and could not be tracked in cycle 090712
Hermine	10L	090600	090800	9	9	
Igor	11L	090812	092112	53	53	
Julia	12L	091212	092012	33	32	LO 092012
Karl	13L	091418	091806	15	15	
Lisa	14L	092100	092612	23	23	
Matthew	15L	092318	092612	12	12	
Nicole	16L	092812 093006	092918 093006	6 1	6 1	
Otto	17L	100606	101006	17	17	
Paula	18L	101118	101506	15	15	
Richard	19L	102100	102612	23	23	
Shary	20L	102900	103018	8	8	

Thomas	21L	102918	110718	37	37	
Total				426	402	
2010 Pacific						
Blas	03E	061712	062118	18	18	
Celia	04E	061906	062900	40	39	LO 062900
Darby	05E	062300	062900	25	25	
Six	06E	071500 071618	071606 071618	6 1	6 1	
Estelle	07E	080600	081012	19	19	
Eight	08E	082012	082112	5	5	
Frank	09E	082118	082812	28	28	
Ten	10E	090306	090406	5	5	
Eleven	11E	090400	090406	2	2	Vortex problem 090412
Georgette	12E	092112	092300	7	7	
Total				156	155	
Total 2010				582	557	
Total Atlantic				697	660	
Total Pacific				485	468	
Test Total				1182	1128	

Appendix B: Archives

Output files

- Messages
    - domain_center
    - Tcvital
- geogrid output
    - geo_nmm*
    - namelist.wps
- real output
    - namelist.input
    - fort.65
    - wrfinput_d01
- WRF ghost output
    - ghost_d02_0000-00-00_00:00:00
- WRF analysis output
    - wrfanl_d02_yyyy-mm-dd_hh:00:00
- Vortex relocation output
    - wrfinput_d01
    - wrfghost_d02
- GSI output for parent and ghost domains
    - wrf_inout
    - ${SID}L.${yyyymmddhh}.gsi_cvs[1,2].biascr
    - logs
        - stdout
        - fort.201 through fort.215
- Ocean Initialization output
    - ocean_region_info.txt
    - getsst/mask.gfs.dat
    - getsst/sst.gfs.dat
    - getsst/lonlat.gfs
    - phase4/track
    - logs
        - getsst/getsst.out
        - phase3/phase3.out
        - phase4/phase4.out
- Coupled WRF-POM run input and output
    - RST.final
    - wrfinput_d01
    - wrfbdy_d01
    - wrfanl_d02
    - EL.*
    - GRADS.*
    - OHC.*
    - T.*
    - TXY.*
    - U.*
    - V.*
    - WTSW.*
    - rsl.*
- Postprocessing output

    - WRFPRS*
- Tracker output
    - Short track (12 h forecast) from forecasts at 3-h intervals
        - Combined domain
            - gvt_combined_12hr_3hrly_HNR2_${SID}_${yyyymmddhh}.txt
        - Parent domain
            - gvt_parent_12hr_3hrly_HNR2_${SID}_${yyyymmddhh}.txt
    - Long track (126h forecast) from forecasts at 6-h intervals
        - Combined domain
            - fort.64
    - Long track (126h forecast) from forecasts at 3-h intervals
        - Combined domain
            - gvt_parent_12hr_3hrly_HNR2_${SID}_${yyyymmddhh}.txt

- Graphics Output
    - hwrf_plots/${SID}.${yyyymmddhh}/*gif
- Verification Output
    - nhc_HNR2_${SID}_.${yyyymmddhh}.txt
- logs
    - All files

Appendix C: List of Acronyms

3D-Var – Three dimensional Variational Analysis

ATCF – Automated Tropical Cyclone Forecasting

BC – Boundary Conditions

DTC – developmental Testbed Center

EMC – Environmental Modeling Center

GFDL – Geophysical Fluid Dynamics Laboratory

GFS – Global Forecasting System

GSD – Global Systems Division (of NOAA Earth System Research Laboratory)

GSI – Global Statistical Interpolator

GRIB – Gridded binary data format

HNR2 – HWRF configuration used in this test (stands for HWRF njet R2 code)

HR20 – HWRF configuration similar to HNR2 used in a previous test

HWRF – Hurricane Weather Research and Forecasting

IC – Initial Conditions

MSLP – Mean Sea Level Pressure

MSS – Mass Storage System

NAM Post – North American Model Post-processor

NCEP – National Centers for Environmental Prediction

NHC – National Hurricane Center

NMM – Non-hydrostatic Mesoscale Model

NOAA – National Oceanic and Atmospheric Administration

POM – Princeton Ocean Model

Pre13d GFS – Retrospective runs made with GFS implemented operationally on June 29, 2010

RC – reference Configuration

SID – Storm Identification

WPP – WRF Post-Processor

WPS – WRF Preprocessing System

WRF – Weather Research and Forecasting

yyyymmddhh – Year, month, day and hour of forecast initialization