The Developmental Testbed Center

Community HWRF 2011 Operational Capability Final Report

Point of Contact: Ligia Bernardet

August 30, 2011

Contributors:

The DTC Community HWRF 2011 Operational Capability Test was conducted by the team of the Hurricane Task of the DTC, composed by Shaowu Bao, Ligia Bernardet, Mrinal Biswas, Timothy Brown, Laurie Carson, and Donald Stark. This test was designed in collaboration with Vijay Tallapragada of NOAA/NCEP/EMC, who provided the H21A tracks for verification and comparison against HNR2. All acronyms are defined in Appendix C.

Index

1. Executive summary

- The DTC conducted its second extensive test of a HWRF configuration (HD33), demonstrating that a robust testing environment, functionally-similar to EMC's, is available.
- Over 600 HWRF runs for the Eastern North Pacific and North Atlantic basins for the 2010 season were conducted in order to establish a benchmark of the community code (HD33) and to compare the forecasts against a counterpart set produced at EMC (H21A).
- Track errors for HD33 increase linearly with time from near zero at initialization time to 280 nm at the 5-day forecast in both basins.
- 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.
- A negative intensity bias is noted for HD33 in the Pacific basin after the second day of forecasting, while in the Atlantic there is no statistically significant bias.
- The HD33 forecast storm size is larger than the observed one, and continuously grows in size, for the 34-, 50-, and 64-kt wind radii in the Atlantic basin. In the Eastern North Pacific, the forecast size is over predicted at the onset, but decreases with forecast lead time.
- 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.
- While an exact match between the HD33 and H21A forecasts was not expected due to differences in computational platform and a few other minor setup differences noted in Section 3, a large number of statistically significant differences in track, intensity, and structure were found between the two sets.
- Diagnostic investigations conducted after the test revealed that the differences were caused by a coding error in the convective parameterization. This bug behaved differently in different computational platforms. After correcting this bug, a small sample of forecasts was rerun and indicated that HD33 and H21A results were much closer.
- 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

This report describes a test and evaluation exercise conducted by the Developmental Testbed Center (DTC) for the Hurricane WRF system, known as HWRF (Gopalakrishnan et al. 2011). 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 used by the model developers at NCEP EMC. The configuration tested matches the 2011 Operational HWRF implemented at NCEP on May 15, 2011.

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, UPP, GFDL vortex tracker, GrADS-based graphics, and NHCVx. HWRF is currently designed for use in the North Atlantic and North East 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 Community HWRF 2011 Operational Capability Test was to establish the skill of the community HWRF code for tropical storm forecasting to ascertain that all 2011 operational capabilities have been successfully ported to the community code. This was done through a comparison of the results of the forecasts generated by DTC with the community code (dubbed HD33) against those generated at NOAA NCEP using the H21A configuration of HWRF.

In order to facilitate this comparison, HD33 was run using the prep_hybrid tool to ingest the GFS reforecast pre13j dataset to create initial and boundary conditions for the atmospheric fields. It is recognized that prep_hybrid is not a tool currently supported to the community and that the GFS reforecasts are not easily accessible by the community. Therefore, these runs will not be used to designate a DTC Reference Configuration, which should use community components whenever possible.

It was not expected that HD33 and H21A forecasts would match exactly, but that their bulk verification statistics would be close. There are differences in computing platform, as the HD33 runs were conducted in a Linux cluster and H21A was run n the NCEP operational IBM platform. Additionally, different versions of the tracker were used in the two sets of forecasts. While differences between HD33 and H21A forecasts are relatively small for cold start runs, the cycling nature of HWRF causes an amplification of the differences for the later initializations of a given storm. Table 1 summarizes the differences between HD33 and H21A.

Table 1. Differences between HD33 and H21A

	HD33	H21A
Institution conducting test	DTC	EMC
Computer platform	Linux Cluster *tjet*	NCEP IBMs *CCS*
Source code repository	Community	EMC
Scripts	DTC	EMC
Automation	NOAA GSD Workflow Manager	EMC HWRF History Sequence Manager
I/O format	NetCDF	Binary
UPP	Community UPP Beta v0.5c	EMC UPP 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 is composed of WPS, prep_hybrid, vortex relocation and initialization, GSI, ocean initialization, POM, WRF, coupler, UPP, 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 for prep_hybrid and NHCVx, which are not currently supported to the community. NHCVx was obtained from a DTC in-house code repository. The revisions for all codes are listed below:

- WRF - https://svn-wrf-model.cgd.ucar.edu, revision 4947
- WPS – https://svn-wrf-wps.cgd.ucar.edu, revision 602
- UPP – Beta release v0.5c (revision 75)
- GSI – official release v2.5
- Vortex Relocation and Initialization, prep_hybrid, miscellaneous libraries and tools: https://svn-dtc-hwrf-tne.cgd.ucar.edu, revision 245
- POM and POM initialization – https://svn-dtc-pomtc.cgd.ucar.edu, revision 85

- Coupler - https://svn-dtc-ncep-coupler.cgd.ucar.edu, revision 37
- Tracker - https://svn-dtc-gfdl-vortextracker.cgd.ucar.edu, revision 53
- NHCVx - https://svn-dtc-nhcvx.cgd.ucar.edu, revision 28

The scripts were obtained from the DTC in-house repository at https://svn-dtc-hwrf-tne.cgd.ucar.edu, revision 158.

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 covers a 75x75° area with 0.18° (approximately 27 km) horizontal grid spacing. There are a total of 216 x 432 grid points in the parent grid. The nest covers a 5.4 x 5.4° area with 0.06° (approximately 9 km) grid spacing. There are a total of 60 x 100 grid points in the nest. Both parent and nest use 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, vary from run to run and are 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 depends on the location of the storm at the initialization time and on the 72-h NHC forecast for the storm location. Those parameters define whether the East Atlantic or United domain of the POM are used. Both POM domains cover an area from 10.0°N to 47.5°N in latitude, with 225 latitudinal grid points. The East Atlantic POM domain ranges from 60.0° W to 30.0° W longitude with 157 longitudinal grid points, while the United domain ranges from 98.5° W to 50.0° W with 254 longitudinal grid points. Both domains have horizontal grid spacing of approximately 18 km in both the latitudinal and longitudinal directions. The POM uses 23 vertical levels and employs the terrain-following sigma coordinate system.

Additional intermediate domains are used for the atmospheric model during the vortex relocation and initialization procedures (see Bao et al. 2011), 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.   Initial and Boundary Conditions

Initial Conditions were based on pre13j GFS analysis. Pre13j GFS refers to the retrospective runs of the GFS implemented operationally on May 9, 2011. 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 applies a vortex relocation procedure as described in Bao et al. (2011) and Gopalakrishnan et al. (2011). 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 removes the vortex from the GFS analysis and substitutes it with the vortex from the previous HWRF forecast, after correcting it using the observed location and intensity. When a previous HWRF forecast is not present, the GFS vortex is 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*.

For storms classified as *deep* by the NHC at the time of model initialization, the HWRF initialization is updated using GSI. The data supplied to GSI consists 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 are employed because of quality and availability of the datasets. No data (except satellite radiances) is assimilated in the inner core of the storm, that is, the GSI modifications to the HWRF initialization are only applied to the storm environment (outside 150 km radius from the storm center).

6

d. Forecast periods

Forecasts were initialized every 6 hours for the storms listed in Table 5 (Appendix A) and run out to 126 hours. 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.

e. Physics suite

The physics suite configuration (Gopalakrishnan et al. 2011) is described in Table 2. The convective parameterization is applied in both the parent and nest domains with momentum mixing activated in both.

Table 2. Physics Suite for HD33 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 (84)

f. 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 was done using the Lagrangian scheme.

g. Post-processing and vortex tracking

The unipost program within UPP 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 UPP, the post-processed parent and nest domains were both 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 were be horizontally interpolated to a high-resolution standard latitude-longitude grid with similar domain to the nest in order to generate graphics.

Tracking was performed 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 was done with three-hourly postprocessed files.

h. Model verification

The characteristics of the forecast storm (location, intensity, structure) as contained in the HD33 and H21A ATCF files produced by the tracker were compared against the Best Track using the NHCVx. The HD33 ATCF files were produced by the DTC as part of this test, while the H21A ATCF files were supplied by EMC. The NHCVx was run separately for each case, at 6-hourly forecast lead times, out to 126 h, in order to generate a distribution of errors.

A R-statistical language script was run separately on an homogenous sample of the HD33 and H21A datasets to aggregate the errors and to create summary metrics including the mean and median of track error, along-and across 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 in HD33 forecasts (worst forecasts) were identified.

Additionally, pairwise differences (HD33-H21A) of track error, along-and across track error, intensity error and absolute intensity error were computed and aggregated with a R-statistical language script. Ninety-five percent confidence intervals were computed to determine if there is a statistically significant (SS) difference between the two configurations.

i. Graphics

Graphics were generated using GrADS scripts originally developed at EMC. Graphics include 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 10-m winds on the nest
- Zonal cross sections of zonal and meridional wind on the nest

- Meridional cross section of zonal wind on the nest

All graphics are displayed on the DTC Testing and evaluation website (http://verif.rap.ucar.edu/eval/hwrf_hd33_h21a/).

j. 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_pre13j/*yyyymmdd*, where *yyyymmdd* is the year, month and day of the forecast initialization.

The TCVitals, the A- and B-decks (containing H21A 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_HD33_run_archive/dataset.tar.bz2.

The output can be found at /arch3/jet/projects/dtc-hurr/HWRF_HD33_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_HD33_run_archive/tracks_HD33_H21A.tar.bz2 contains all the track files for HD33 and H21A.

A file with the output from the NHC Vx for all HD33 and H21A cases can be found at /arch3/jet/projects/dtc-hurr/HWRF_HD33_run_archive/HWRF_HD33_H21A_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_HD33_run_archive/HD33_H21A_Rscript_and_output.tar.bz2

All source codes and executables are in /arch3/jet/projects/dtc-hurr/HWRF_HD33_run_archive/TNE_source_files.tar.bz2.

All logs are in /arch3/jet/projects/dtc-hurr/HWRF_HD33_run_archive/HWRF_HD33_logs.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 was done 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 were 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 was written summarizing the results and conclusions from this test

7. Results

For brevity, this report gives a summary of the most important results. A comprehensive set of verification figures is available at http://verif.rap.ucar.edu/eval/hwrf_hd33_h21a/verify/.

a. North Atlantic basin

The mean of the track errors for HD33 and H21A indicates that the errors grow in time from near zero at the initialization time to approximately 250 nm at the five-day lead time (Fig. 2). The along-track errors (Fig. 3) indicate that both HD33 and H21A forecasts are systematically too slow, especially in the middle of the forecast period. Cross-track errors (Fig. 4) are near zero for the HD33 configuration in the first day of forecast, and increase after that (not SS), indicating a tendency of positioning the storm to the right of the observed one. On the other hand, the H21A configuration displays small SS positive errors on the first day of forecast, which reduce to near zero on the second and third day of the forecast and grow towards the end of the forecast period, but do not become SS.

Figure 2. Mean track error (nm) for HD33 (black) and H21A (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.
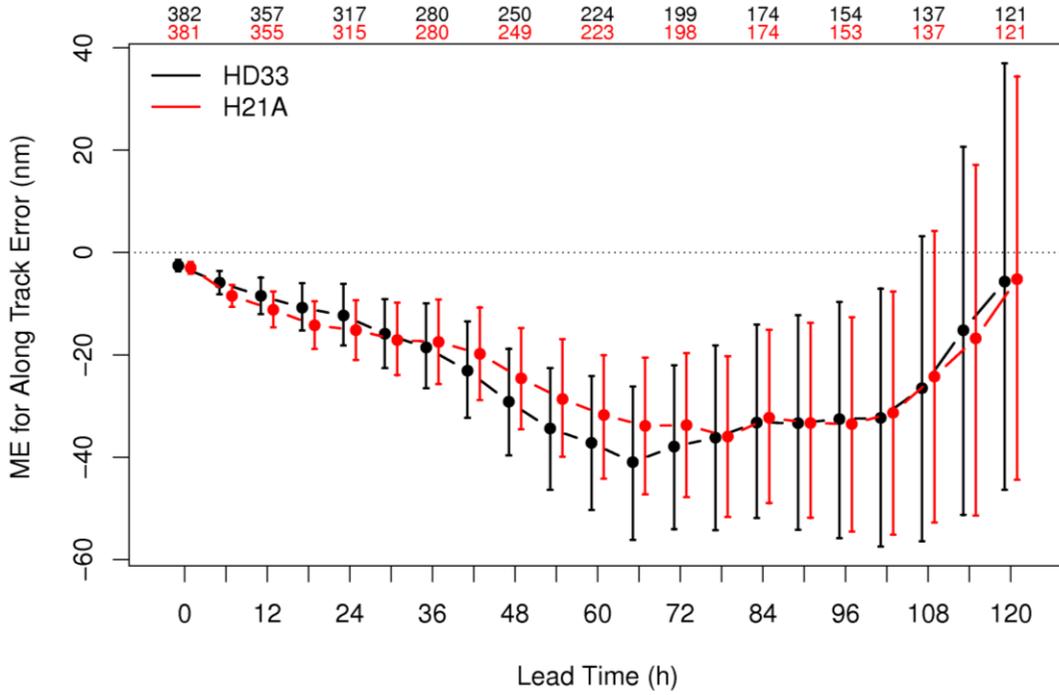
Figure 3. Same as Fig. 2, except for along track error (nm). Positive numbers indicate forecasts are too fast and negative numbers indicate they are too slow.

Figure 4. Same as Fig. 2, except for cross track error (nm). Positive numbers indicate forecasts deviate to the right of the observed path, while negative numbers indicate forecasts deviate to the left.

There are seven SS differences in the median of track errors between HD33 and H21A, occurring between the third and fifth day of the forecast (Fig. 5). Differences in absolute cross-track errors are SS for lead times 48- through 66 h, while differences are SS for lead times 72, 90, and 96 h for absolute along-track error. All differences favors the H21A configuration (Table 3).
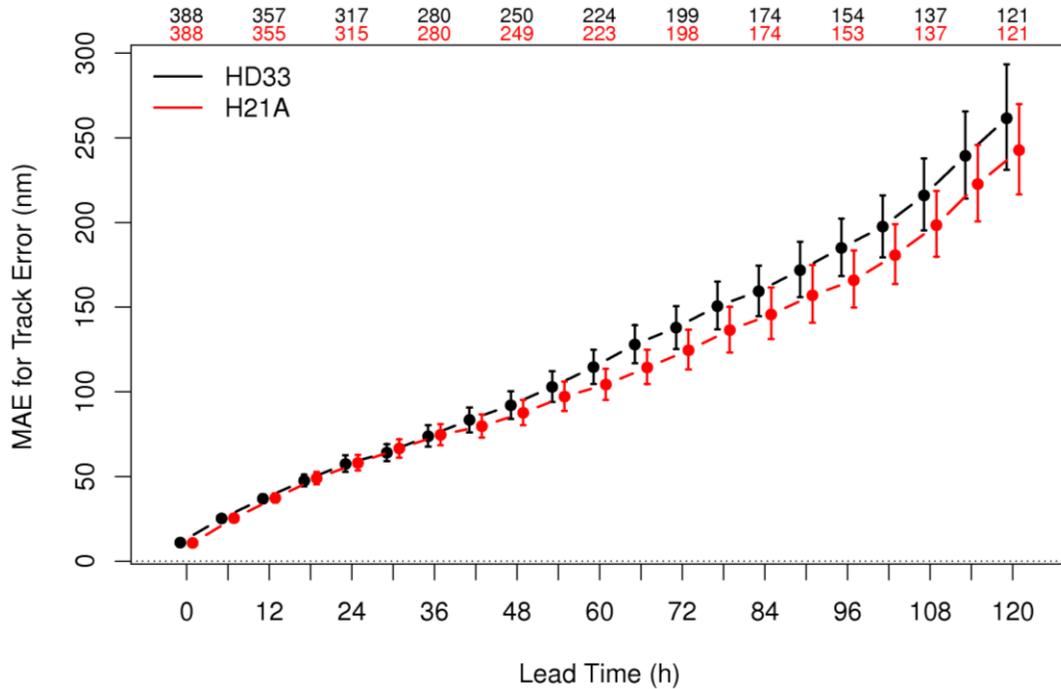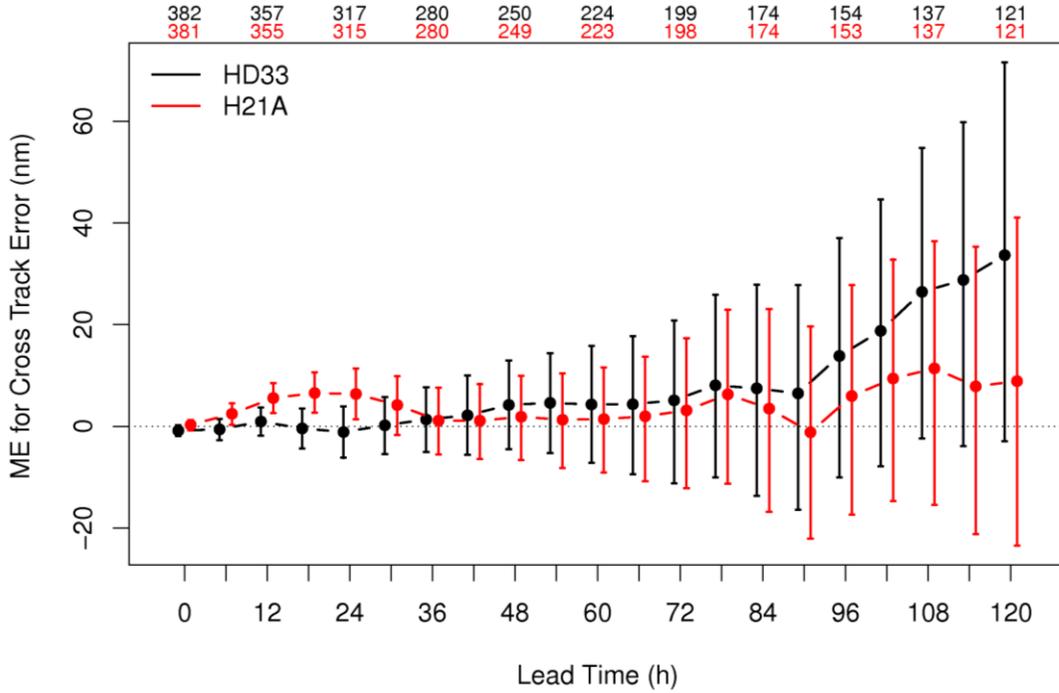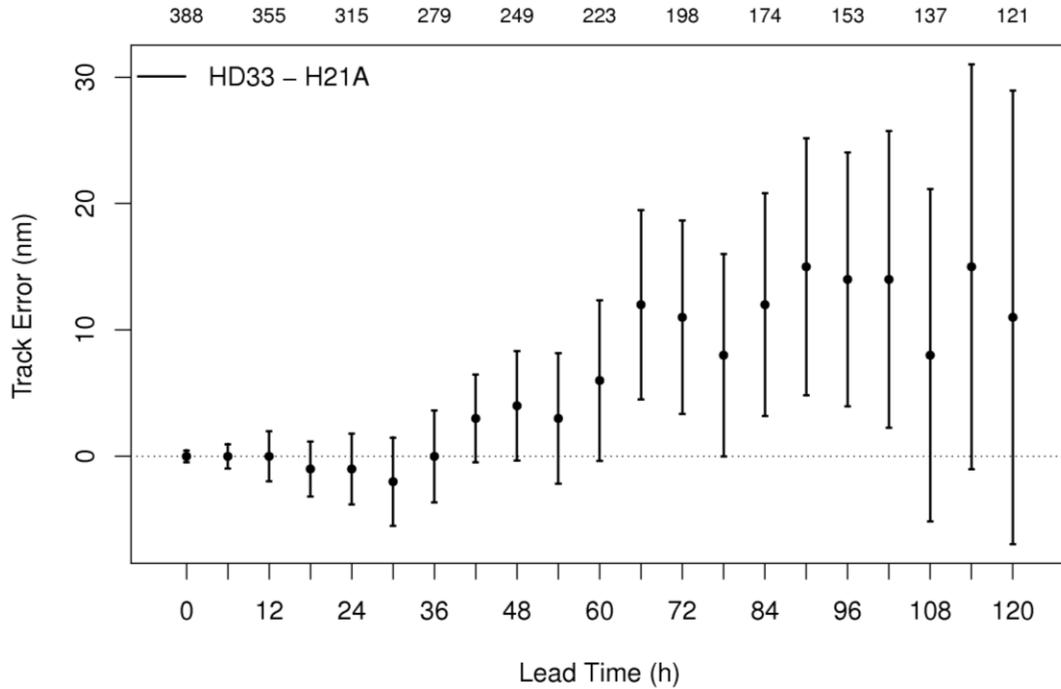

Figure 5. Median of the pairwise difference in track error (nm) between HD33 and H21A as a function of forecast lead time for all cases in the Atlantic basin. The 95% confidence intervals are also displayed. Positive (negative) values indicate H21A (HD33) superior performance. The sample size is listed above the graphic.

The mean of the absolute intensity errors for HD33 and H21A displays a sharp growth in the first 6 h of forecast, from near 3 to 8 kt. From 6 to 48 h, the error grows more slowly to about 14 kt, and grows slowly after that (Fig. 6). The mean of the intensity errors for HD33 is slightly negative in the first two days of forecast, is slightly positive in the third and half of the fourth day of forecast, and returns to negative numbers after that. The bias for HD33 remains between ±2 kt, and is only SS at the initial time). On the other hand, H21A is initialized with only -0.5 kt of bias, but small positive bias are seen in the first day of forecast (up to 1 kt and not SS). On days two and beyond, H21A displays negative biases that get worse in time, reaching -7 kt by the end of the forecast period (Fig. 7). These differences between HD33 and H21A result in nine SS differences between the two configurations. At the initialization, H21A is favored, while later in the forecast HD33 is favored (Fig. 8).
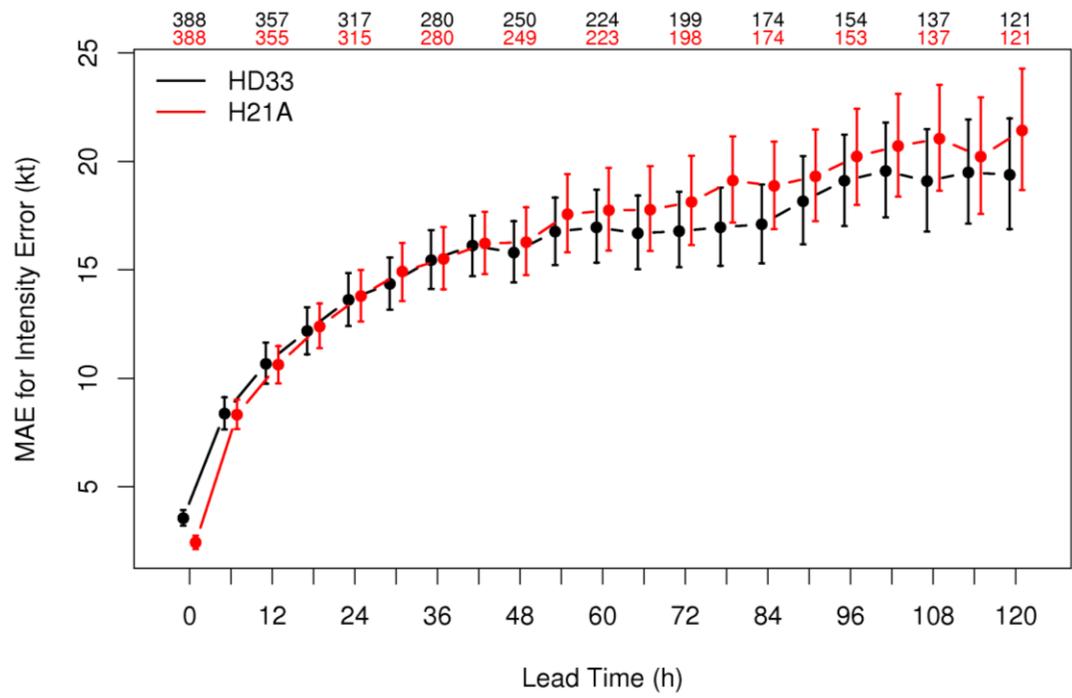
Table 3. Statistical significances in the mean of various errors as a function of forecast lead time for all cases in the Atlantic basin. The lines correspond to track error (Tk), absolute cross-track error (CrTk), absolute along-track error (AlTk), absolute intensity error (Wd), radii of the 34, 50, 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 HD33 or H21A configurations are indicated as "D" or "A", 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	-	-	-	-	-	-	-	-	-	-	-	A	A	A	A	A	A	A	-	-	-
CrTk	-	-	-	-	-	-	-	-	A	A	A	A	-	-	-	-	-	-	-	-	-
AlTk	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	A	A	-	-	-	-
Wd	A	-	-	-	-	-	-	-	D	-	-	-	D	D	-	D	-	D	D	D	D
NE34	-	D	D	D	D	D	D	D	-	-	-	-	-	-	-	-	-	-	-	-	-
NE50	-	D	D	D	D	-	-	-	-	-	-	A	A	A	A	A	A	A	A	A	A
NE64	-	-	D	D	-	D	-	-	-	-	-	A	A	A	A	A	A	A	A	A	A
NW34	-	-	D	D	D	D	-	-	-	-	-	-	-	A	A	A	A	A	A	A	A
NW50	-	-	-	-	-	-	-	-	-	-	-	-	-	A	A	A	A	A	A	A	A
NW64	-	-	D	D	-	D	-	-	A	A	-	A	A	A	A	A	A	A	A	A	A
SE34	D	D	D	D	D	D	-	D	D	-	-	-	-	-	-	-	-	-	D	-	-
SE50	-	D	D	D	D	-	-	-	-	-	-	-	-	-	A	A	A	A	A	A	A
SE64	-	-	D	-	D	D	-	-	-	-	-	-	A	-	A	A	A	A	A	A	A
SW34	-	-	D	D	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	A	A
SW50	-	-	-	-	-	-	-	-	-	-	-	A	-	A	A	-	A	A	-	A	A
SW64	D	-	D	-	-	-	-	-	-	-	-	-	A	A	A	A	A	A	A	A	A

Figure 6. Same as Fig. 2, except for absolute intensity error (kt).
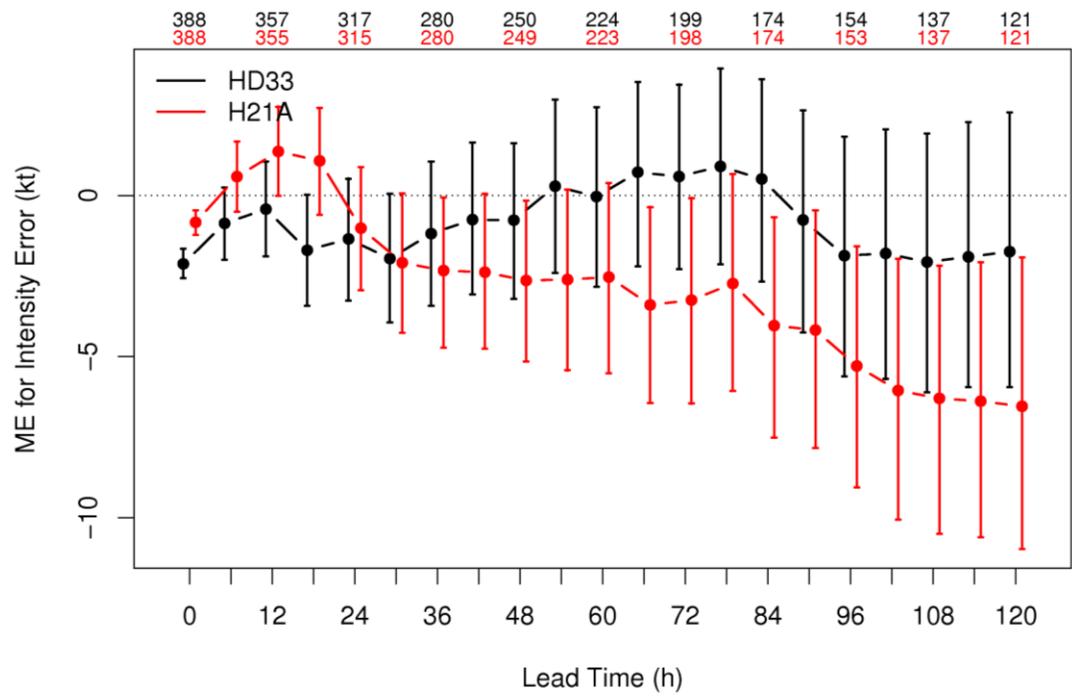

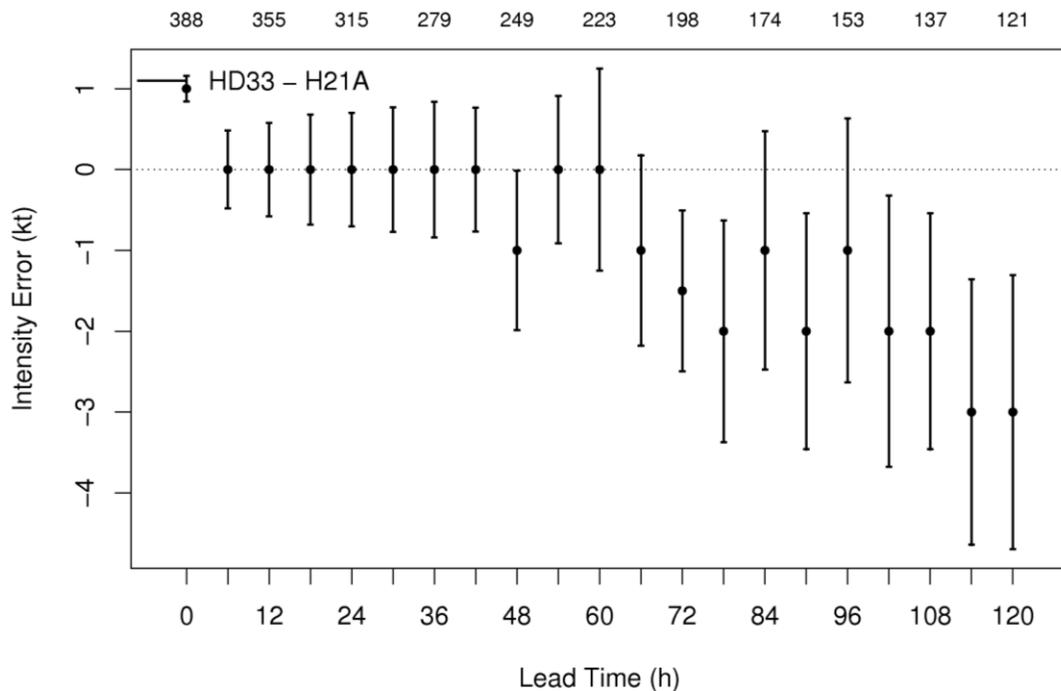Figure 7. Same as Fig. 2, except for intensity error (kt).

Figure 8. Same as Fig. 3, except for absolute intensity error (kt).

The verification of storm structure indicates that both model configurations consistently overestimate storm size, but the mean errors for the wind radii vary substantially among quadrants, wind thresholds, and model configurations (not shown). The larger errors are seen for the 34 kt threshold, for which the mean surpasses 65 nm. Individual errors can be much higher, surpassing 300 nm. Errors for the HD33 model configuration typically grow linearly in time, indicating that the forecast storm expands in time (one exception is the NE quadrant for the 34-kt wind threshold, for which errors have small variation in time). On the other hand, H21A errors can be almost constant in time (such as for the 50-kt threshold in the SE quadrant), peak in the middle of the forecast (such as for the 34-kt threshold in the NE quadrant, which peaks at the 18-h lead time), or have multiple maxima (such as the 34-kt threshold in the SE quadrant, which peaks at the 12, 78, and 114-h lead times).

In general, the H21A configuration overestimates storm size more than the HD33 configuration in the first one or two days of forecast. Since the HD33 storms grow in size during the forecast, the size overestimation is worse for HD33 towards the end of the forecast period. This behavior is clearly depicted in Table 3, which shows the SS differences between HD33 and H21A. Several SS differences favoring HD33 are present in the beginning of the forecast, while numerous SS differences favoring H21A can be seen in the four and five-day forecast lead time.

Figure 9 displays the outliers in track forecasting for the Atlantic Basin. The worst forecasts for the HD33 configuration are annotated. In the first two days, the worst

forecasts are from Nicole, while Lisa is responsible for several outliers on days 3 and 4. Finally, the worst 5-day forecasts are from Danielle and Richard.

The worst storms on average are not necessarily the same as the ones responsible for the largest outliers. When the error is computed individually for each storm (see http://verif.rap.ucar.edu/eval/hwrf_hd33_h21a), the largest mean track errors are for Richard (700 nm in 5 days) Lisa (500 nm in 5 days), Colin (400 nm in 5 days), Nicole (exceeds 400 nm in 42 h) and Otto (more than 300 nm in 78 h).
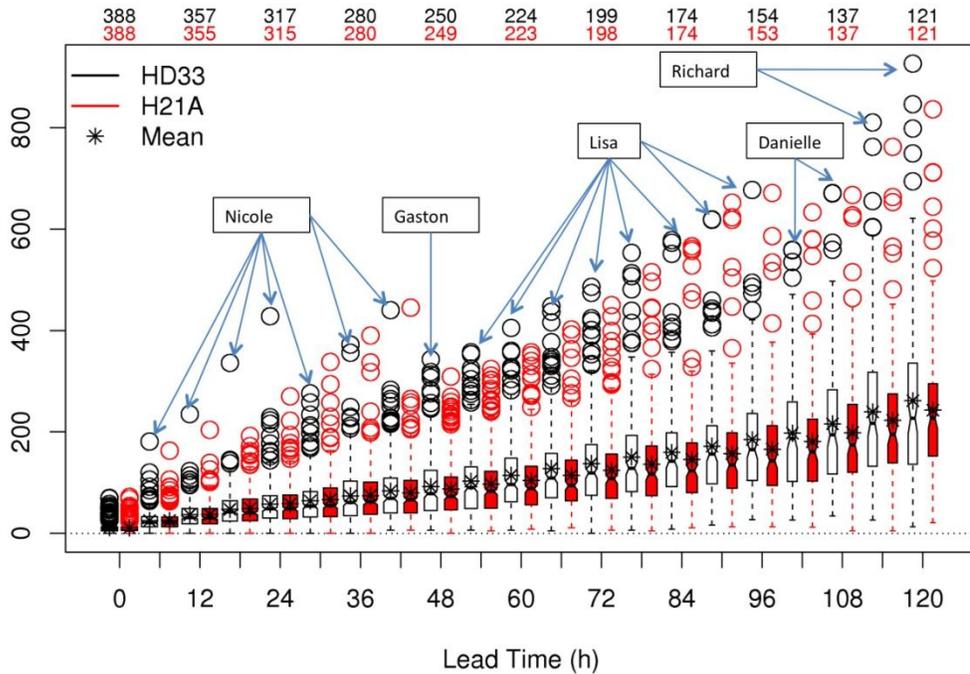

Figure 9. Boxplots of mean track errors for the HD33 (black) and H21A (red) configurations. The median is the waist of the plot and the 95% confidence intervals are the notches. The bottom and top of the boxplots denote the 25th and 75th percentiles, respectively. Outliers are represented as circles. A star represents the mean.

The boxplots and outliers for absolute intensity can be seen in Fig. 10. The distributions are broad with a large number of outliers, indicating that there is a strong variability in the forecast skill from run to run. The worst forecasts are spread over a variety of storms. The largest outliers come from Karl, Paula and Igor on the first two days of forecast. Danielle and Julia are responsible for the errors on the third day of forecast. On the fourth day, the storm with the largest outlier is Karl, and finally on day 5, largest errors come from Fiona and Richard.

While on average the intensity errors tend to grow sharply in the first few forecast hours and subsequently become saturated, errors in individual storms can have maxima in the middle of the forecast period. Therefore, individual storms do not have the worst forecast throughout the entire forecast period. Intensity errors for Paula and Julia peak at 24 and 60 h, respectively, making them the worst storms on average for those lead times. Conversely, errors for Colin and Richard grow sharply in the last few hours of forecast, making them the worst storm on average for the 5-

day forecast. Large intensity errors can be caused by track errors when the observed storm moves inland and the forecast storm does not, or vice versa. In the future, it would be interesting to identify the worst intensity forecasts that are not caused by interactions with land.
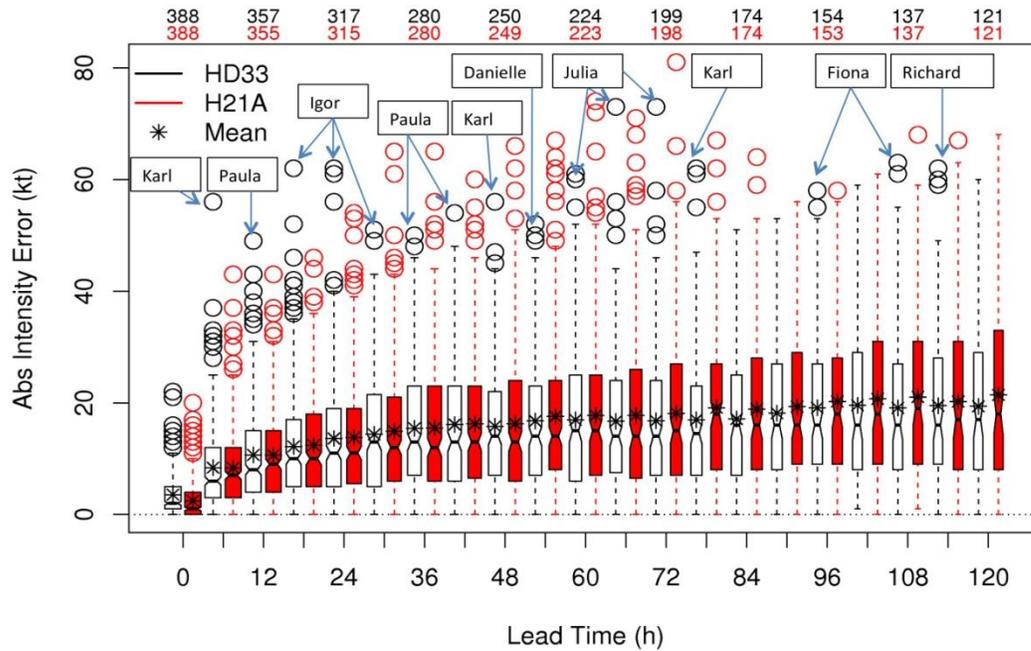

Figure 10. Same as Fig, 9, but for absolute intensity.

b. Eastern North Pacific basin

The magnitude of mean track errors is similar in the North Atlantic and Eastern North Pacific basins for the HD33 configuration. However, for the H21A configuration, track errors in the Eastern North Pacific basin are larger than in the Atlantic one, increasing to over 350 nm for the 5-day forecast (Fig. 11).

Along track errors on the Eastern North Pacific basin are near zero in the first two days of forecast and negative after that, indicating that storm motion is too slow. The errors become progressively larger (more negative) as time goes on, and surpass -100 and -200 nm for the HD33 and H21A configurations, respectively.

Pacific cross-track errors grow in time from near zero at the beginning of the forecast period to a maximum of 80 and 130 nm for the HD33 and H21A model configurations, respectively. The positive errors indicate that forecasts from both configurations deviate to the right of the observed storm.

While the average track errors are similar in the two basins, both the along- and cross- track errors are much larger in the Eastern North Atlantic basin than in the

North Atlantic one. This indicates that the Eastern North Pacific track errors are dominated by bias (too slow and too much to the right), while the North Pacific track errors have less bias.

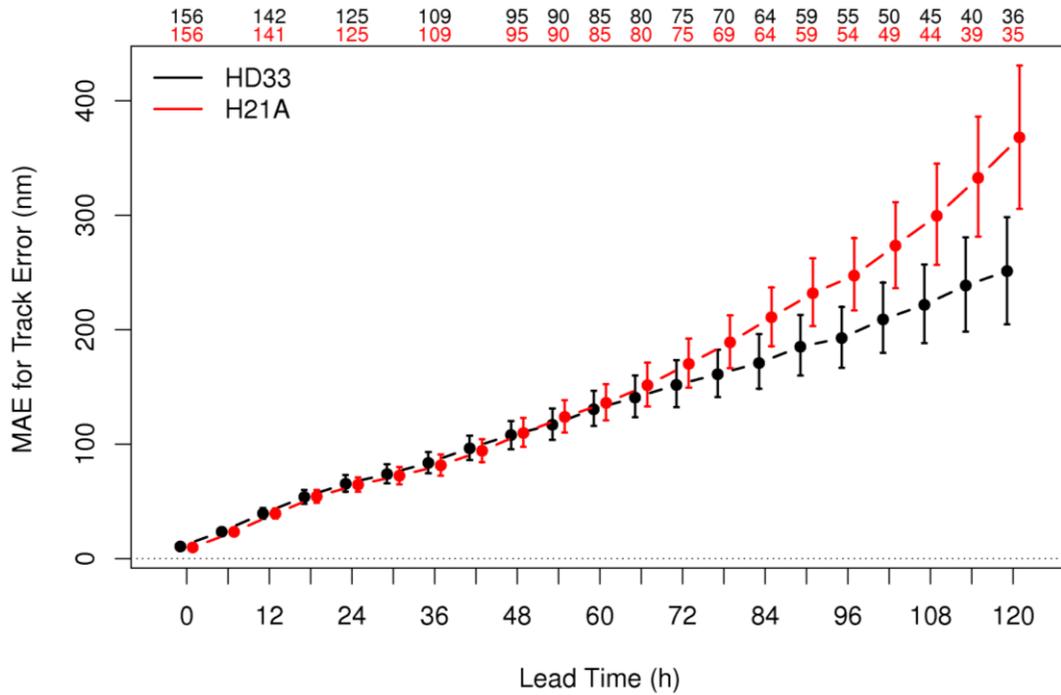**Figure 11. Mean track error (nm) for HD33 (black) and H21A (red) as a function of forecast lead time for all cases in the North Pacific basin. The 95% confidence intervals are also displayed. The sample size is listed above the graphic.**
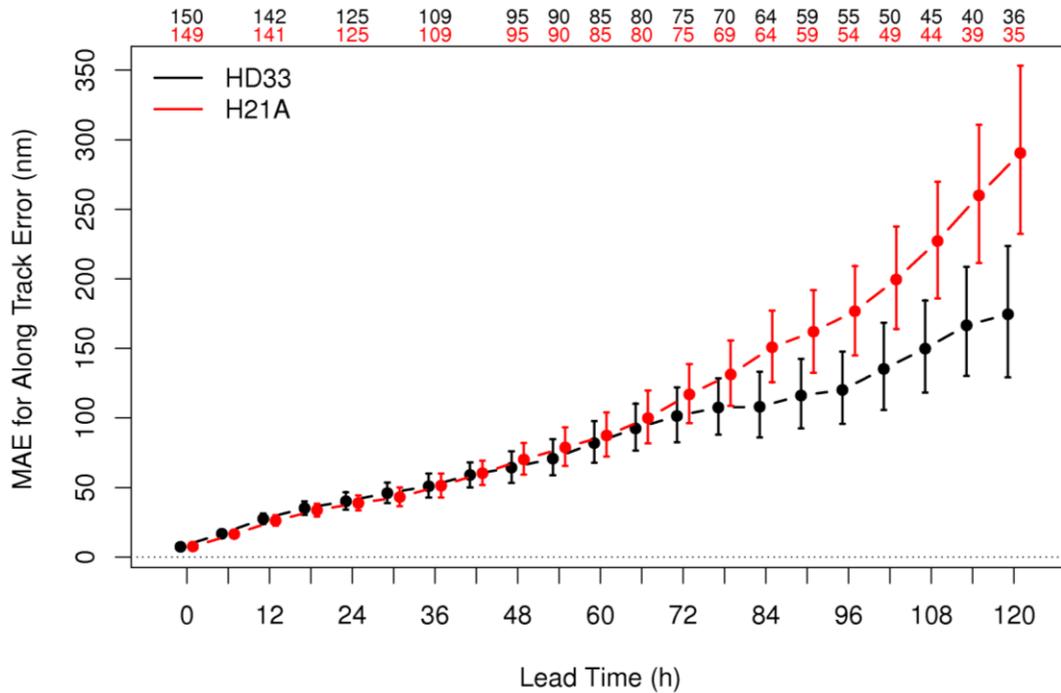

Figure 12. Same as Fig. 11, except for along-track error (nm). Positive numbers indicate forecasts are too fast and negative numbers indicate they are too slow.
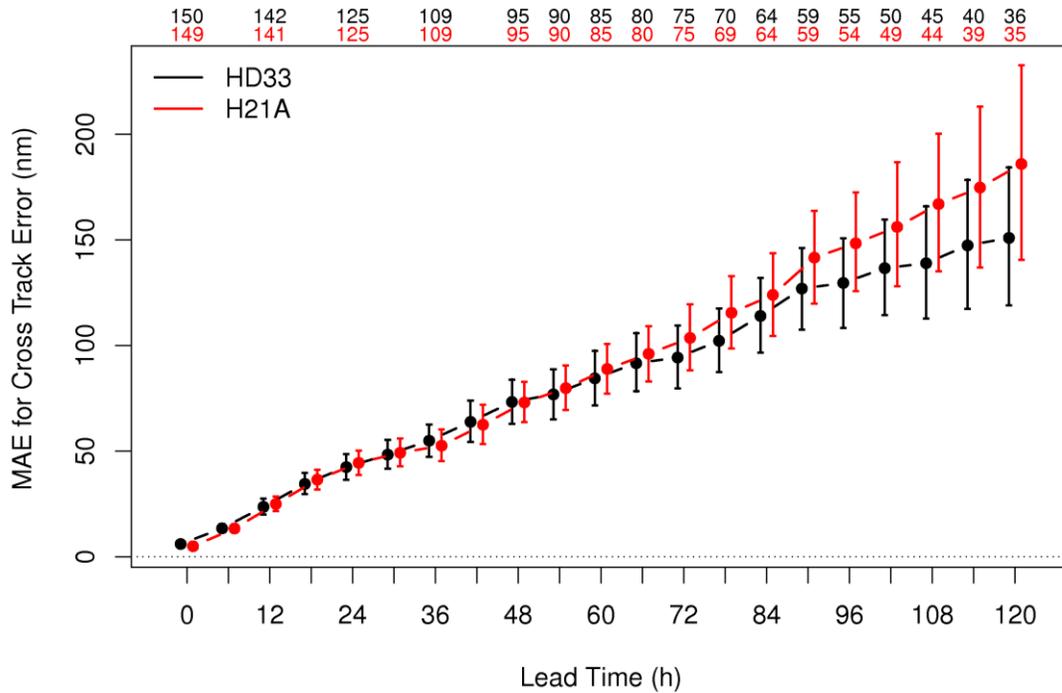
Figure 13. Same as Fig. 11, except for cross-track error (nm). Positive numbers indicate forecasts deviate to the right of the observed path, while negative numbers indicate forecasts deviate to the left.

Statistically significant differences in track error between HD33 and H21A appear at the 36- and 42-h lead times favoring H21A and at 60- and 72- through 120-h favoring HD33 (Fig. 14 and Table 4). These results differ sharply from those obtained for the North Atlantic basin, in which all seven SS differences favors the H21A configuration. Several SS differences occur in along- and cross-track error. The vast majority occurs in the last two days of forecast and favors the HD33 configuration (Table 4).

The mean absolute intensity error is similar to the Atlantic basin, but errors grow more steadily throughout the forecast period (Fig. 15). At the 5-day forecast, errors exceed 25 kt, and are therefore larger than their North Atlantic counterparts. The mean intensity errors (Fig. 16) indicate that the errors are small but SS negative at the initial time. The underestimation of intensity is reduced in the first day of forecast, and actually gives way to overestimation at 6- and 12-h for the H21A configuration. Later in the forecast period, the mean intensity errors become smaller and turn negative again, with the intensity underestimation becoming worse as lead time increases.

Statistically significant differences in storm intensity absolute errors (Fig. 17) only occur at the initial time (favoring H21A), and at the 6- and 12-h lead times, both favoring the HD33 configuration. The magnitude of the differences does not exceed 1 kt.

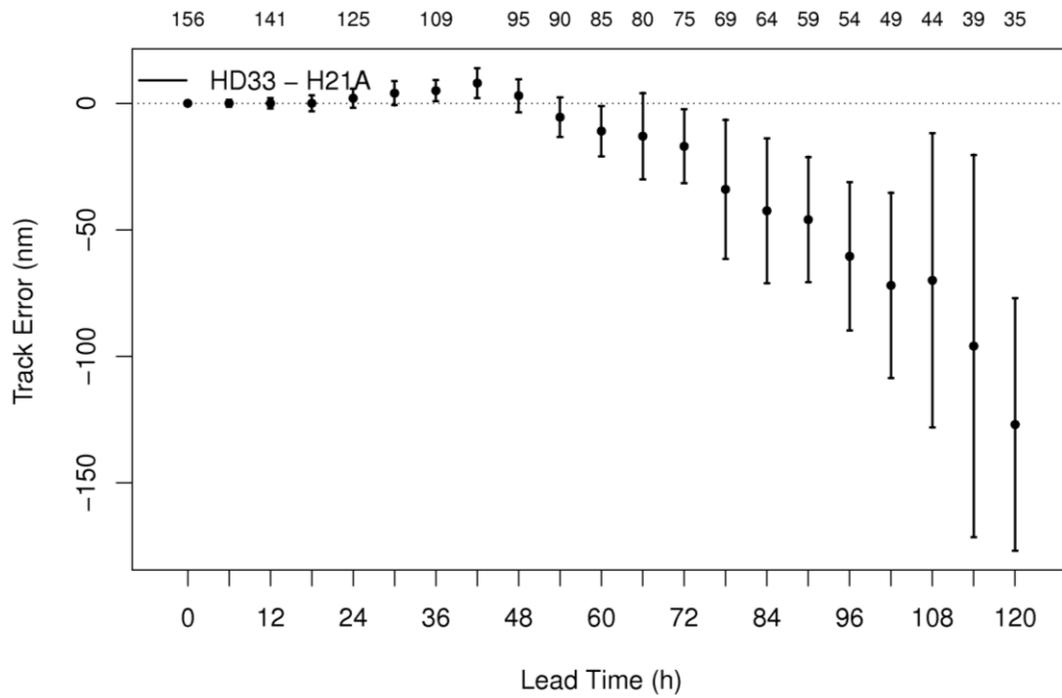**Figure 14. Median of the pairwise difference in track error (nm) between HD33 and H21A as a function of forecast lead time for all cases in the Eastern North Pacific basin. The 95% confidence intervals are also displayed. Positive (negative) values indicate H21A (HD33) superior performance. The sample size is listed above the graphic.**
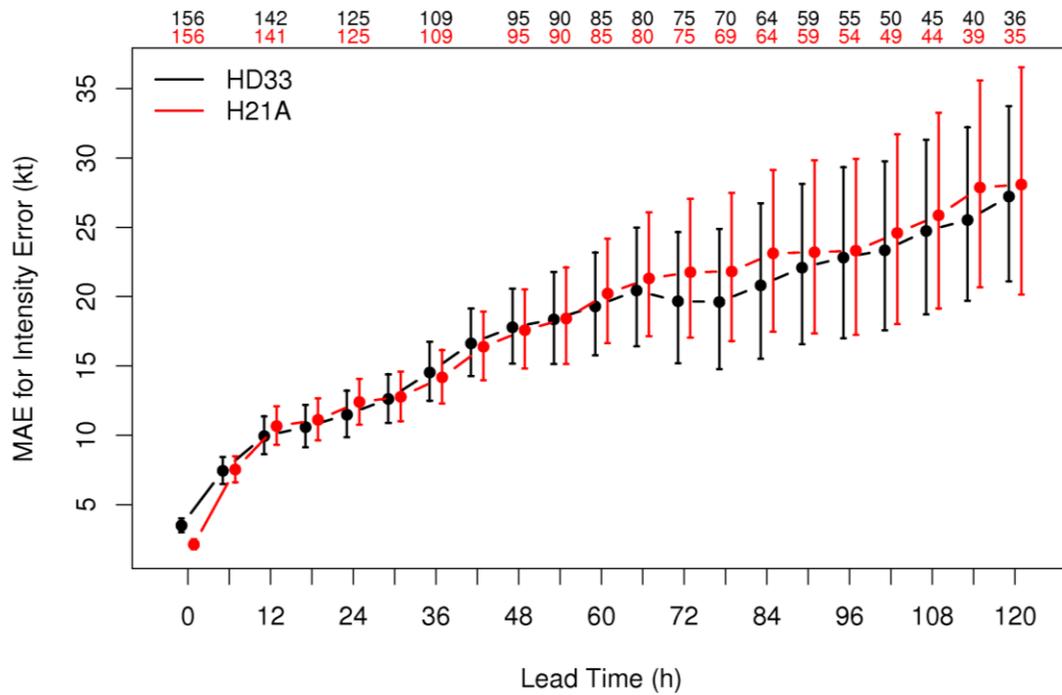

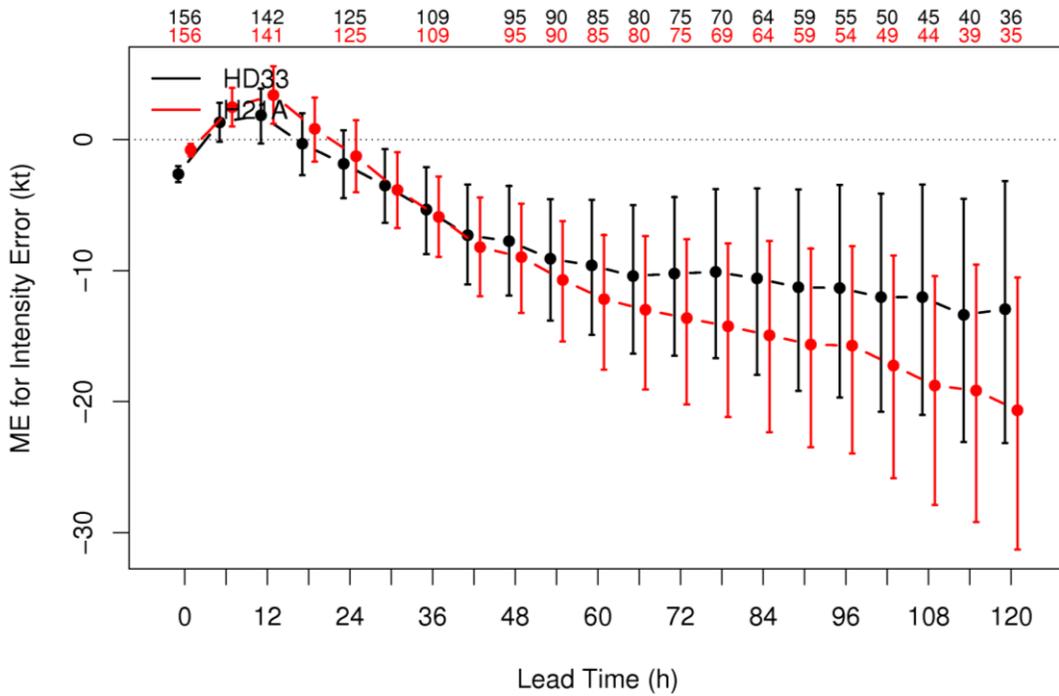**Figure 15. Same as Fig. 11, but for absolute intensity error (kt).**

Table 4. Same as Table 3, except for 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	-	-	-	-	-	-	A	A	-	-	D	-	D	D	D	D	D	D	D	D	D
CrTk	-	-	D	-	-	-	-	-	-	-	-	-	D	D	-	D	-	D	D	D	-
AlTk	-	-	-	-	-	A	-	-	-	-	-	-	-	-	D	-	D	D	-	D	D
Wd	A	D	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NE34	-	D	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D
NE50	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	-
NE64	-	-	-	-	-	D	D	-	-	-	-	-	A	-	-	-	-	-	-	-	-
NW34	-	D	-	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	D	D
NW50	-	-	-	-	-	-	-	-	-	-	-	-	D	-	D	-	-	-	-	D	D
NW64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	-	-	-	-
SE34	D	D	D	D	D	D	-	D	-	-	-	-	-	-	-	-	A	A	-	-	D
SE50	-	D	D	D	D	D	D	D	-	-	-	-	A	A	A	A	A	-	-	-	-
SE64	-	-	-	D	D	D	D	D	-	-	-	-	-	-	-	-	-	-	-	-	-
SW34	-	D	D	D	D	D	D	D	D	-	D	-	-	-	-	-	-	A	-	A	A
SW50	D	D	-	D	-	-	-	-	-	-	-	-	A	-	-	A	A	-	-	-	-
SW64	-	-	-	-	-	-	D	D	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 16. Same as Fig. 11, but for intensity errors (kt).
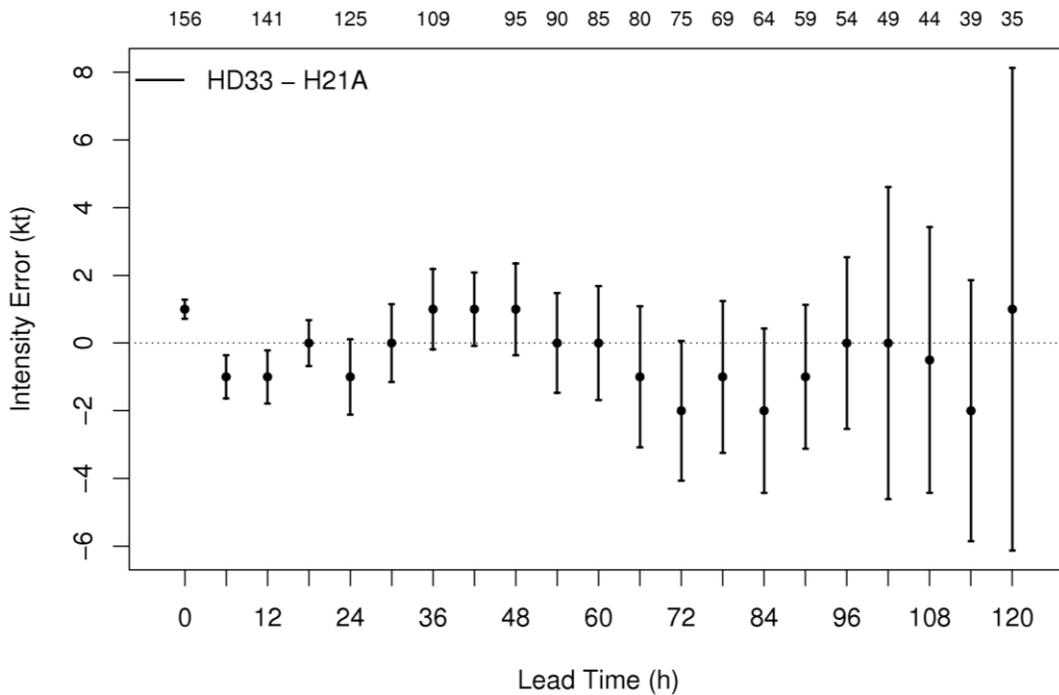

Figure 17. Same as Fig 14, but for absolute intensity errors.

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 in the first half of the forecast period, but becomes smaller than the observed ones towards the end of the forecast. This result indicates that the forecast storms tend to decrease with forecast

lead time. This behavior is markedly different from the North Atlantic one, where the forecast storms become larger in time. It should be noted that the sample size is much smaller in the Pacific than in the Atlantic, especially for stronger storms, so these results need to be interpreted with caution.

Differences in structure between HD33 and H21A are summarized in Table 4. Numerous SS differences occur between HD33 and H21A in the radii of the various wind thresholds. Most of them occur in the first two days of forecasting, and indicate that the overforecasting of storm size is worst for the H21A configuration. Some SS favoring H21A occur later in the forecast, particularly for the 50-kt threshold in the SE quadrant.

Figure 18 displays the outliers in track forecasting for the Eastern North Pacific Basin for the HD33 configuration. The worst forecasts in the first day are from Ten (Td10). Estelle has the worst errors in the second day of forecast, while Darby is responsible for the largest errors in days 3-5. These three storms also appear as the ones with largest mean error of the Eastern North Pacific basin for the HD33 configuration.

The outliers for absolute intensity can be seen in Fig. 19. The worst individual forecasts are for Celia, which is also the storm with the worst average absolute intensity forecast.
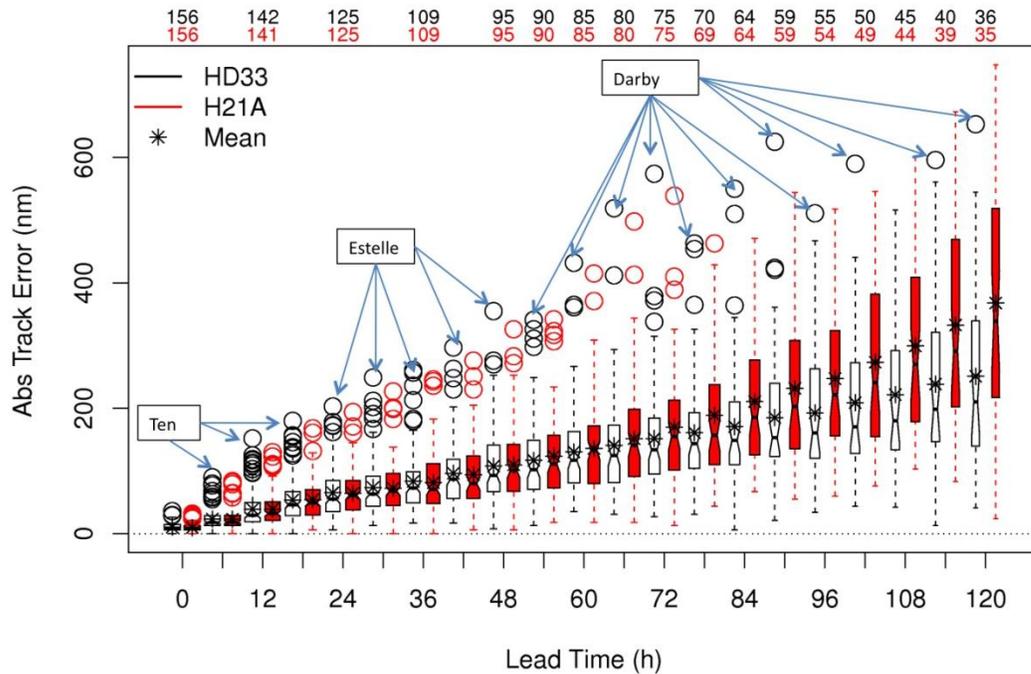

Figure 18. Boxplots of mean track errors for the HD33 (black) and H21A (red) configurations for the Eastern North Pacific Basin. The median is the waist of the plot and the 95% confidence intervals are the notches. The bottom and top of the boxplots denote the 25 th and 75 th percentiles, respectively. Outliers are represented as circles. A star represents the mean.
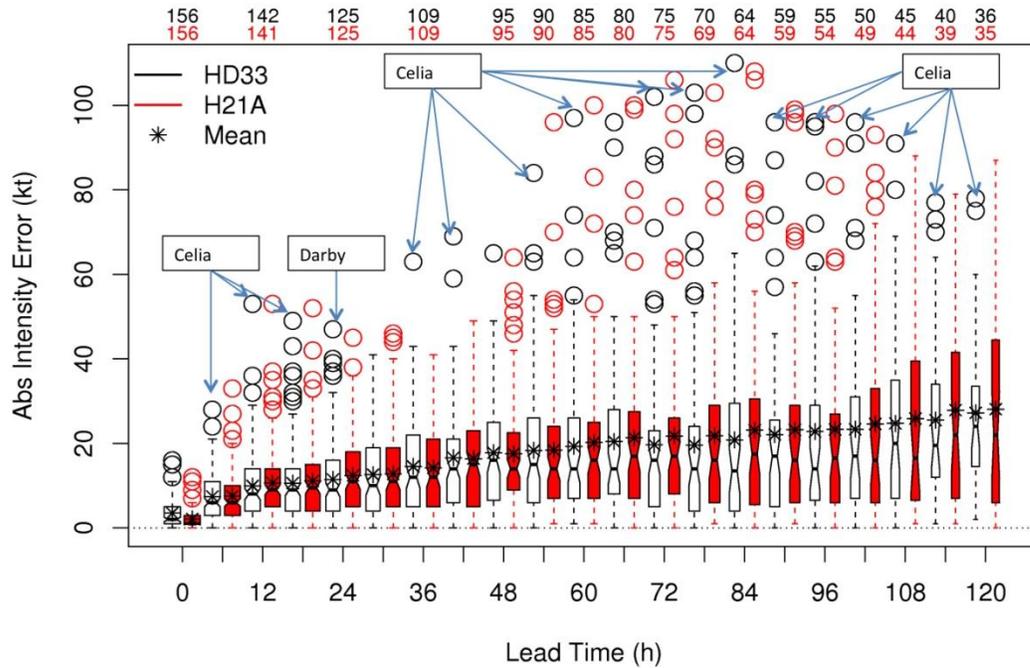
Figure 19. Same as Fig. 14 but for the absolute intensity errors.

8. Interpretation and conclusions

This experiment was the second extensive test of HWRF configurations conducted by the DTC. The HD33 runs were very robust. However, in three occasions the external tracker lost the storm, and therefore the files for continued vortex cycling were not produced (Karl 09/18 at 00 UTC, Nicole 09/29 at 12 UTC, and Tomas 11/03 12 UTC).

The verification results obtained indicated significant differences between the HD33 and H21A. Tracks from the H21A configuration were better in the Atlantic, while their HD33 counterparts were better in the Pacific. Several SS differences in wind intensity favoring the HD33 configuration were noted in both basins. These differences were unexpected because H21A and HD33 are very similar configurations, which differ only in the points noted in Table 1.

Subsequent investigation of the source of the differences was conducted focusing on Darby. This storm was chosen because large differences between H21A and HD33 appear already in the cold start. As part of the investigation, the H21A setup was run on the same computational platform used to run HD33, and the results were dubbed H21J. As seen in Fig. 20, the forecasts by H21A and H21J are very different, indicating that computational platform plays a significant role in determining the forecast results.
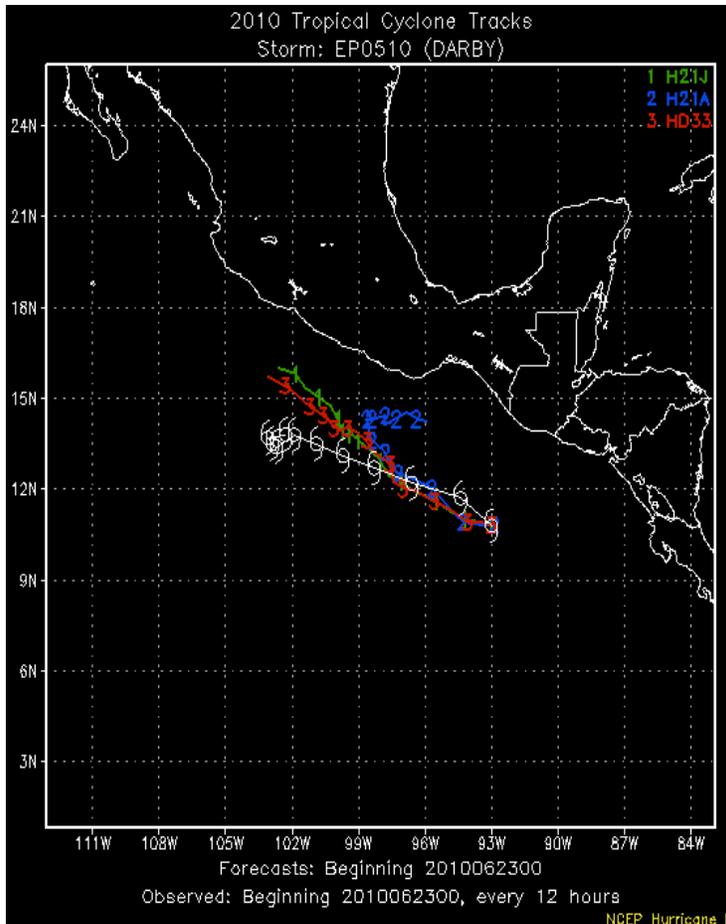
Figure 20. Forecast tracks by the H21J (green), H21A (blue), and HD33 (red) model configurations for Darby initialized on 06/23/2010 at 00 UTC. The Best Track (white) is also displayed.

This extreme sensitivity of the forecasts to the computational platform, to the extent of showing up as SS differences in bulk statistical results, had not been seen in previous HWRF tests conducted by the DTC. Additional diagnostics revealed the presence of an uninitialized variable in the cumulus parameterization code. After fixing this bug, the results for the Darby test case became similar between the two platforms. This bug fix was then incorporated in the community HWRF code, is part of the August 04, 2011 HWRF V3.3a release, and will be applied in all subsequent HWRF tests. This bug was also fixed in the 2011 operational HWRF in the beginning of August, 2011.

Even though the entire test was not rerun after this bug was corrected, the case study investigation that followed the test indicated that, with the fix, the community code is producing results similar to the code housed in the EMC repository. This community code, with the bug fixed, will constitute a baseline for the development of the 2012 operational HWRF, and will be used in subsequent testing aimed at establishing a new HWRF Reference Configuration.

9. References

Bao, S., R. Yablonsky, D. Stark, and L. Bernardet, 2011. HWRF Users' Guide V3.3a. Developmental Testbed Center, 88pp.

Gopalakrishnan, S., Q. Liu, T. Marchok, D. Sheinin, N. Surgi, R. Tuleya, R. Yablonsky, and X. Zhang, 2011: Hurricane Weather and Research and Forecasting (HWRF) Model: 2011 scientific documentation. L. Bernardet, Ed., 75 pp.

Acknowledgements

Appendix A: Inventory

Table 5. Inventory for HD33 Test. Columns on the table refer to the storm name, storm number, number of cases in the test plan, 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.

2010 Atlantic					# nhcvx files	# filled nhcvx files	Notes
Alex	01L	22	062600	070106	22	22	
Two	02L	2	070806	070812	2	2	
Bonnie	03L	10	072212	072418	10	10	
Collin	04L	24	080218	080812	24	17	LO 080400-080512
Five	05L	4	081100	081118	4	4	
		1	081218	081218	1	0	LO
		2	081306	081312	2	0	LO
		9	081506	081706	9	0	LO
Danielle	06L	37	082118	083018	37	37	
Earl	07L	41	082512	090412	41	41	
Fiona	08L	16	083100	090318	16	16	
Gaston	09L	7	090112	090300	7	5	LO 090218, 090300 LO 090312, 090318, 090400, 090500-090800
		17	090312	090712	17	3	
Hermine	10L	9	090600	090800	9	9	
Igor	11L	53	090812	092112	53	53	
Julia	12L	33	091212	092012	33	32	LO 092012
Karl	13L	15	091418	091806	14	14	091806 failed because storm is weak and 3-hrly tracker for 12-h forecast of previous cycle failed
Lisa	14L	23	092100	092612	23	23	
Matthew	15L	12	092318	092612	12	12	
Nicole	16L	6	092812	092918	6	5	092918 failed because storm is weak and 3-hrly tracker for 12-h forecast of previous cycle failed
		1	093006	093006	1	1	
Otto	17L	17	100606	101006	17	17	
Paula	18L	15	101118	101506	15	15	
Richard	19L	23	102100	102612	23	23	
Shary	20L	8	102900	103018	8	8	
Tomas	21L	37	102918	110718	19	19	110318 failed because storm is weak and 3-hrly tracker for 12-h forecast of previous cycle failed
Total Atlantic		444			425	388	

2010 Pacific						
Blas	03E	18	061712	062118	18	18	
Celia	04E	40	061906	062900	40	39	LO 062900
Darby	05E	25	062300	062900	25	25	
Six	06E	6 1	071500 071618	071606 071618	6 1	6 1	
Estelle	07E	19	080600	081012	19	19	
Eight	08E	5	082012	082112	5	5	
Frank	09E	28	082118	082812	28	28	
Ten	10E	5	090306	090406	5	5	
Eleven	11E	3	090400	090412	3	3	
Georgette	12E	7	092112	092300	7	7	
Total Pacific		157			157	156	
Total Test		601			582	544	

Appendix B: Files to be archived in the MSS

- 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 domain
  - wrf_inout
  - ${SID}L.${yyyymmddhh}.gsi_cvs1.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
    - sharpn/sharpn.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
  - Long track (126h forecast) from forecasts at 6-h intervals

- Combined domain
  - fort.64
- Graphics Output
  - hwrf_plots/${SID}.${yyyymmddhh}/*gif

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

HD33 – HWRF configuration used in this test (stands for HWRF DTC v3.3)

H21A – HWRF configuration similar to HD33 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

NHCVx - National Hurricane Center verification package

NMM – Non-hydrostatic Mesoscale Model

NOAA – National Oceanic and Atmospheric Administration

POM – Princeton Ocean Model

Pre13j GFS – Retrospective runs made with GFS

SID – Storm Identification

SS – Statistically significant

UPP – Unified Post-Processor

WPS – WRF Preprocessing System

WRF – Weather Research and Forecasting

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