NOAA Hurricane Forecast Improvement Project

High-Resolution Hurricane Forecast Test

Final report

Developmental Testbed Center

09/30/2009

HRH Executive Summary

The HFIP High-Resolution Hurricane Test was conducted by the Developmental Testbed Center in the period March – 2008 through September – 2009 with the goal of assessing the impacts of using higher horizontal resolution in hurricane numerical forecasting. The plan for this test was developed jointly by several segments of the community, including specialists in hurricanes, numerical modeling, and forecast verification. Additionally, six independent modeling groups participated in this effort. The results of this test are summarized below for each participant modeling group.

1. **HWRF-X model contributed by AOML:** Increasing horizontal resolution in the AOML model reduced track and intensity errors for the short to intermediate lead times and improved the frequency of forecasted RI events. However, higher resolution degraded wind radii forecasts and led to overforecasting of RI episodes.
2. **AHW model contributed by MMM:** Increasing the horizontal resolution of the MMM model reduced the track error for longer lead times. While higher resolution did not have a significant impact on the intensity errors, it did improve the forecast's ability to capture the observed frequency of RI events.
3. **COAMPS-TC model contributed by NRL:** The higher resolution NRL configuration had a positive impact on intensity for a few lead times and improved RI and RW forecasts, but degraded the track and wind radii forecasts.
4. **ARW model contributed by PSU:** No conclusions can be reached for the PSU model due to the small sample size and large case-to-case and lead time to lead time variability in performance.
5. **GFDL model contributed by URI:** The higher-resolution URI model did not substantially improve the errors in track or intensity and it increased the wind radii errors.
6. **UW-NMS model contributed by UWM:** Increasing the resolution of the UWM model had a positive impact on intensity for several lead times and the frequency of forecasted RI events, but degraded track and wind radii forecasts and increased the FAR for RI events.

The use of higher resolution in the participating models did not lead to an overall benefit in tropical cyclone forecasting as measured by the metrics used in this study. Improvement was noted for some metrics, lead times and models but the majority of results showed no difference in using high resolution and a few, notably wind radii, presented consistent degradation when using high resolution.

The largest improvement in intensity with least degradation in other areas was seen for the AOML and MMM models. It is possible that the benefits of high-resolution were not fully realized in the participating models due to limitations, such as physics suites that are not appropriate for high-resolution, lack of a coupled ocean model, initialization techniques, or the nature of the numerics themselves (e.g., GFDL model is hydrostatic). Additionally, it is possible that the resolutions used in the test are not high-enough to resolve small scale structures such as updrafts and meso-vortices that may need to be represented in order to improve intensity forecasting.

We recommend diagnostic studies be conducted for a small sample of cases to determine if processes important to intensification are missing in the forecast. Once those are identified and addressed by the use of alternative physics suites and/or initialization techniques, new comprehensive tests can be conducted and it is possible that the benefits of high-resolution may be realized.

Index

List of Figures

Figure 6.2.5: Same as Fig. 6.1.5, except for MMM.

Figure 6.2.6: Same as Fig. 6.1.6, except for MMM.

Figure 6.3.1: Same as Fig. 6.1.1 except for NRL configurations.

Figure 6.3.2: Distributions of the pairwise differences between the track errors for the high- and low-resolution configuration of NRL with respect to lead time.

Figure 6.3.3: Absolute intensity error distributions with respect to lead time for the high- and low-resolution NRL model configurations.

Figure 6.3.4: Same as Fig. 6.1.3 except for NRL configurations.

Figure 6.3.5: Same as Fig. 6.1.4 except for NRL configurations.

Figure 6.3.6: Same as Fig. 6.1.5, except for NRL.

Figure 6.3.7: Same as Fig. 6.1.6, except for NRL and a) contains all forecasts and b) only those for which both low- and high-resolution runs are complete.

Figure 6.4.1: Scatter plots of PSU track error with respect to lead time for a) low- and intermediate-resolutions configurations, and b) low- and high-resolution configurations.

Figure 6.4.2: Same as Fig. 6.1.3 except for PSU configurations (top panel – low- and intermediate-resolutions, bottom panel – low- and high-resolutions)

Figure 6.4.3: Same as Fig. 6.1.4 except for PSU configurations.

Figure 6.5.1: Same as Fig. 6.1.1 except for URI configurations.

Figure 6.5.2: Intensity error distributions with respect to lead time for the high- and low-resolution URI model configurations.

Figure 6.5.3: Distributions of the pairwise differences between the absolute intensity errors for the high- and low-resolution configuration of URI with respect to lead time.

Figure 6.5.4: Same as Fig. 6.1.3 except for URI configurations.

Figure 6.5.5: Same as Fig. 6.1.4 except for URI configurations.

Figure 6.5.6: Same as Fig. 6.1.5, except for URI.

Figure 6.5.6: Same as Fig. 6.1.6, except for URI.

Figure 6.6.1: Same as Fig. 6.1.1, except for UWM low- and intermediate-resolutions.

Figure 6.6.2: Same as Fig. 6.5.2, except for UWM low- and intermediate-resolutions.

Figure 6.6.3: Same as Fig. 6.1.3, except for UWM low- and intermediate-resolutions.

Figure 6.6.4: Same as Fig. 6.1.4, except for UWM low- and intermediate-resolutions.

1. Introduction

Tropical cyclones are a serious concern for the nation, causing significant risk to life, property and economic vitality. The National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) has a mission of issuing tropical cyclone forecasts and warnings, aimed at protecting life and property and enhancing the national economy. In the last 10 years, the errors in hurricane track forecasts have been reduced by about 50% through improved model guidance, enhanced observations, and forecaster expertise. However, little progress has been made during this period toward reducing forecasted intensity errors.

To address this shortcoming, NOAA established the Hurricane Forecast Improvement Project (HFIP) in 2007. HFIP is a 10-year plan to improve one to five day tropical cyclone forecasts, with a focus on rapid intensity change (for more details, see http://www.nrc.noaa.gov/HFIP%20Draft%20Plan.html). The HFIP plan details a variety of approaches to improve hurricane forecasting, including research and development to provide: (1) an observing strategy analysis capability for hurricanes, (2) an improved understanding of hurricane intensity change, and (3) an advanced hurricane numerical modeling system. Recent research suggests that prediction models with grid spacing less than 1 km in the inner core of the hurricane may provide a substantial improvement in intensity forecasts (Powers and Davis 2002, Hendricks et al. 2004, Yau et al. 2004, Braun et al. 2006, Chen et al. 2007, Davis et al. 2008, Rotunno et al. 2009). The 2008-09 staging of the High Resolution Hurricane (HRH) Test focused on quantifying the impact of increased horizontal resolution in numerical models on hurricane intensity forecasts. The primary goal of this test was an evaluation of the effect of increasing horizontal resolution within a given model across a variety of storms with different intensity, location and structure. A secondary goal was to provide a data set that can be used to explore the potential value of a multi-model ensemble for improving hurricane forecasts.

The Developmental Testbed Center (DTC) and the HFIP Team hosted a workshop at the National Hurricane Center (NHC) in Miami, FL, 11-12 March 2008. Experts on hurricanes, numerical modeling and model evaluation met for two days to discuss the strategy for this test (see Appendix A for list of the workshop participants). The test plan reflecting the consensus reached during this workshop on the framework for this testing effort is posted at http://www.dtcenter.org/plots/hrh_test/HIRES_HFS_Test_Plan.pdf. The HRH Test Plan put forth minimum guidelines for the retrospective forecasts that would be evaluated. To isolate the impact of high resolution, each modeling group was required to submit retrospective forecasts for at least two horizontal resolutions where the low-resolution forecasts were not influenced by a higher resolution nested domain.

A second HRH workshop was held at the NHC in Miami, FL, 7-8 May 2009 to discuss preliminary results based on the retrospective forecasts submitted prior to the workshop and potential future collaborations (see Appendix A for a list of participants). Presentations from the second workshop are posted at: http://www.dtcenter.org/plots/hrh_test/workshop2009/. This report briefly describes the retrospective cases selected for this test, the model configurations used to generate retrospective forecasts, the evaluation system used to generate objective verification statistics, and the results for the complete data set.

2. HRH Participants

2.1 DTC Evaluation Team

The DTC Evaluation Team was tasked with assembling a state of the art hurricane verification system, assembling the required data sets, providing an Output Module to the modeling groups for generating the required Gridded Binary 1 (GRIB1) format output, collecting gridded output from the participating modeling groups, processing the gridded output to generate objective verification statistics focused on quantifying the impact of resolution, and preparing the final project report. This team was composed of the following DTC staff:

Scientists			Software Engineers
Ligia Bernardet	Jamie Wolff	Huiling Yuan	Christopher Harrop
Louisa Nance	Edward Szoke	Edward Tollerud	John Halley Gotway
Barbara Brown	Tara Jensen	Tressa Fowler	Laurie Carson
Shaowu Bao			Paula McCaslin

2.2 Modeling groups

Six modeling groups participated in the HRH test (see Table 2.1). The range of model resolutions and number of test cases for which forecasts were delivered varied depending on the resources each group had available to commit to the project. The models used for HRH included three configurations of the Weather Research and Forecasting (WRF) model, the operational Geophysical Fluid Dynamics Laboratory (GFDL) model, the Navy's tropical cyclone model, and the University of Wisconsin-Madison (UWM) Non-hydrostatic Modeling System (UW-NMS).

Table 2.1: Modeling groups that participated in HRH and the models used to generate retrospective forecasts.

Institution (Contact)	Model	Grid Spacings		
		Low	Mid	High
NOAA /Atlantic Oceanic and Meteorological Laboratory (S. Gopalakrishnan)	HWRF-X	9 km	3 km	-
National Center for Atmospheric Research / Mesoscale and Microscale Meteorology (Chris Davis)	AHW	12 km	-	1.33 km
Naval Research Laboratory (Melinda Peng)	COAMPS-TC	9 km	3 km	-
Pennsylvania State University (Fuqing Zhang)	WRF-ARW	13.5 km	4.5 km	1.5 km
University of Rhode Island (Isaac Ginis)	GFDL	1/12° (~9 km)	1/18° (~6 km)	-
University of Wisconsin-Madison (Greg Tripoli)	UW-NMS	12 km	3 km	1 km

2.3 Verification Team

The verification team determined the sample size for the cases, the criteria for case selection, the verification metrics for assessing the difference in skill due to changes in model resolution, what tools are available for computing the selected metrics, and finally what observations, analyses, and model output fields were needed to compute the selected metrics. This team was composed of a good mixture of research and operational scientists: Barb Brown (NCAR), James Franklin (NHC), Mike Fiorino (NHC), Mark DeMaria (NOAA/CIRA), and Tim Marchok (GFDL).

2.4 Case Selection Team

The case selection team selected a diverse set of storms and time periods from each of these storms that met the criteria set forth by the verification team. This set of storms features a number of Rapid Intensification (RI) and Rapid Weakening (RW) events. The DeMaria-Kaplan RI criteria defines an RI event as a 30 kt increase in maximum sustained surface wind (MSSW) in a 24-h period, whereas an RW event is defined as a 25 kt decrease in MSSW over water in a 24-h period. Both RI and RW events are restricted to time periods for which the storm is over water. The Case Selection Team also included a research and an operational scientist: Mark DeMaria (NOAA/CIRA) and Jack Beven (NHC).

3 Model Descriptions

This section provides basic descriptions of the model configurations run for HRH. More information and references for each configuration can be found on the model description tab of the HRH website: http://www.dtcenter.org/plots/hrh_test/model/index.php. Gridded data submitted to the DTC for evaluation was interpolated by the modeling groups from the model's native grid to an unstaggered uniform-spaced latitude-longitude projection on standard pressure surfaces. For larger storms, a domain of at least 15 x 15 degrees is needed for assessing wind radii quantities. The innermost nest for all model configurations used for this project did not meet this minimum size criterion. Modeling groups used one of two methods to meet the needs for evaluating wind radii: 1) submit the appropriate coarser domain from the nested run to cover 15 x 15 degrees, or 2) generate a blended grid based on the high-resolution nest and its parent to cover 15 x 15 degrees. The domains for which gridded data were delivered to the DTC for evaluation are indicated in the tables below by the inclusion of the assigned model identification in parentheses.

Atlantic Oceanographic and Meteorological Laboratory

NOAA's Atlantic Oceanographic and Meteorological Laboratory (AOML) ran a variant of the WRF Nonhydrostatic Mesoscale Model (NMM) v3.0 (Janjic et al. 2001) referred to as HWRF-X (Experimental Hurricane WRF - Gopalakrishnan et al. 2006, Yeh et al. 2008). The WRF NMM uses the Arakawa E grid on a rotated latitude-longitude projection with a hybrid sigma-pressure vertical coordinate. For the HRH Test, the model was run with one fixed coarse mesh domain and one moving, two-way interactive nested domain. The basic properties of the two model configurations used for HRH are summarized in Tables 3.1-3.3. Both configurations were run with 42 full vertical levels and a model top at 50 hPa.

Table 3.1: AOML configuration used for 9 km runs

	DOMAIN 1 (AOM6)	DOMAIN 2 (AOM1)
Grid movement	Static	Moving
Horizontal grid spacing	27 km	9 km
Horizontal dimensions	320 x 320	88 x 88

Table 3.2: AOML Configuration used for 3 km runs

	DOMAIN 1 (AOM5)	DOMAIN 2 (AOM2)
Grid movement	Static	Moving
Horizontal grid spacing	9 km	3 km
Horizontal dimensions	960 x 960	262 x 262

Table 3.3: Physical parameterizations used in AOML configurations

PARAMETERIZATION	SCHEME USED
Cumulus	Simplified Arakawa-Schubert (27 and 9 km only)
Microphysics	Ferrier
Radiation	RRTM (longwave) / Dudhia (shortwave)
Planetary Boundary Layer	MRF
Surface Layer	Tuleya and Kurihara (1978)
Land Surface	Noah Land Surface Model

Since the inner nest for both configurations was too small to evaluate wind radii, domains 1 and 2 for each configuration were submitted to complete the evaluation. Domain 1 was only used to evaluate the wind radii errors.

Atmospheric and sea surface temperature (SST) fields for the retrospective forecasts were initialized with fields from the GFDL model, whereas the land surface model was initialized with fields from the Global Forecast System (GFS) model. Lateral boundary conditions were obtained from the GFS forecast output on a 1-deg grid at 6-hr intervals. These configurations did not include coupling to an ocean model and the SST field was constant throughout the forecasts.

Mesoscale and Microscale Meteorology Division of NCAR

The Mesoscale and Microscale Meteorology (MMM) division of the National Center for Atmospheric Research ran a configuration of the WRF ARW v3.0.1.1 (Skamarock et al. 2008) referred to as the Advanced Hurricane WRF (AHW) model. The WRF ARW uses the Arakawa C grid with a terrain-following mass vertical coordinate. For this test, the ARW was configured to use the Lambert-Conformal projection with one static domain and two moving, two-way interactive nested domains. The basic properties of these two model configurations are summarized in Tables 3.4-3.6. Both configurations were run with 36 full vertical levels and a model top at 20 hPa.

Table 3.4: MMM configuration used for 12 km runs

	DOMAIN 1 (MMM1)
Grid movement	Static
Horizontal grid spacing	12 km
Horizontal dimensions	469 x 424

Table 3.5: MMM Configuration used for 1.33 km runs

	DOMAIN 1	DOMAIN 2 (MMM4)	DOMAIN 3 (MMM3)
Grid movement	Static	Moving	Moving
Horizontal grid spacing	12 km	4 km	1.33 km
Horizontal dimensions	469 x 424	202 x 202	241 x 241

Table 3.6: Physical parameterizations used in MMM configurations

PARAMETERIZATION	SCHEME USED
Cumulus	New Kain Fritsch (12 km only)
Microphysics	WSM5
Radiation	RRTM (longwave) / Dudhia (shortwave)
Planetary Boundary Layer	YSU
Surface Layer	Monin-Obukov
Land Surface	5-layer thermal diffusion soil model

Since the innermost nest for the high-resolution configuration was too small to evaluate wind radii, both domains 2 and 3 for the high-resolution configuration were submitted to complete the evaluation. During the analysis phase of this project, it was discovered that domain 2 of the high-resolution configuration did not meet the minimum size requirements for evaluating the wind radii parameters. Spurious values for the wind field along the boundary of this domain also interfered with the proper function of the vortex tracker program. Hence, no wind radii errors will be presented for the forecasts submitted by MMM.

Atmospheric fields for the retrospective forecasts were initialized with fields obtained from the Ensemble Kalman Filter (EnKF) method in a 6-hour cycling mode, with WRF ARW V2.2 on a 36-km grid, assimilating surface pressure, rawinsonde (including G-IV dropsondes), Aircraft Communications Addressing and Reporting System (ACARS), cloud motion vectors and tropical cyclone best track data each six hours. Six-hour forecasts on the lateral boundaries were taken from the one-degree six-hour GFS forecast valid at the appropriate time. The ensemble was initialized roughly two days prior to being classified as a depression by adding balanced perturbations from the WRF-Variational Data Assimilation System (WRF-Var) to the GFS 36-h forecast valid at the appropriate time. Using an old forecast with high-amplitude perturbations helps the ensemble develop a flow-dependent ensemble more rapidly than starting from short-term forecasts.

The AHW configurations used a one-dimensional ocean mixed-layer model to represent the

ocean. The ocean model initialization was based on daily Hybrid Coordinate Ocean Model (HYCOM) temperature fields with SST from the National Centers for Environmental Prediction (NCEP) real-time global one-half and one-twelfth (when available) degree daily analysis.

Naval Research Laboratory

The Naval Research Laboratory (NRL) ran the Coupled Ocean/Atmosphere Mesoscale Prediction System – Tropical Cyclone (COAMPS-TC), a variant of COAMPS, the Navy operational regional model, dedicated to the prediction of tropical cyclones. COAMPS-TC, which is developed by NRL, consists of data quality control, data assimilation, initialization, a non-hydrostatic atmospheric model and a hydrostatic ocean model (Hodur 1997). The Arakawa C grid is used for both the atmospheric and ocean models. The atmospheric model utilizes the sigma-z vertical coordinate and the ocean model uses the hybrid sigma/z. For the HRH Test, the model was run on a Mercator projection with one fixed coarse mesh domain and either two or three moving, two-way interactive nested domains. Both configurations used 40 vertical levels with a model top at 32 km. The basic properties of the two configurations used for HRH are summarized in the Tables 3.7-3.9.

Table 3.7: NRL configuration used for 9 km runs

	DOMAIN 1	DOMAIN 2	DOMAIN 3 (NRL1)
Grid movement	Static	Moving	Moving
Horizontal grid spacing	81 km	27 km	9 km
Horizontal dimensions	115 x 103	91 x 91	169 x 169

Table 3.8: NRL Configuration used for 3 km runs

	DOMAIN 1	DOMAIN 2	DOMAIN 3 (NRL5)	DOMAIN 4 (NRL2)
Grid movement	Static	Moving	Moving	Moving
Horizontal grid spacing	81 km	27 km	9 km	3 km
Horizontal dimensions	115 x 103	91 x 91	169 x 169	235 x 235

Table 3.9: Physical parameterizations used in NRL configuration

PARAMETERIZATION	SCHEME USED
Cumulus	Kain and Fritsch
Microphysics	Rutledge and Hobbs (1983)
Radiation	Harshvardardet et al. (1987)
Planetary Boundary Layer	Mellor-Yamada, Dissipative heating (Jin et al. 2007)
Surface Layer	Louis (1979), Wang et al. (2002), Sea Spray (Fairall et al. 1996 with recent updates), Level-off drag coefficient for high winds (Donelan et al. 2004)
Land Surface	Force and restore slab land surface model

Since the innermost nest for the high-resolution configuration was too small to evaluate wind radii, both domains 3 and 4 for the high-resolution configuration were submitted to complete the evaluation, with domain 3 only being used to evaluate wind radii.

The Navy Operational Global Atmospheric Prediction System (NOGAPS) fields were used to provide the first guess field for the first forecast for each storm and the COAMPS-TC output from previous 12-hour forecast was used as the first guess for all of the subsequent forecasts. A relocation method was used to place the vortex at the position issued by the Joint Typhoon Warning Center (JTWC) in the first guess field for each simulation. Synthetic observations were then used to enhance the initial vortex structure and the NRL Atmospheric Variational Data Assimilation System (NAVDAS) was used to assimilate the observational data. Boundary conditions were obtained from the NOGAPS forecast output on 1-deg grid at 6-hr intervals. The NRL Coupled Ocean Data Assimilation (NCODA) was used for ocean data assimilation (including altimeter, Special Sensor Microwave Imager –SSM/I, Multi-Channel SST - MCSST, profile and ship data).

Pennsylvania State University

The Pennsylvania State University (PSU) group ran a configuration of the WRF ARW version 2.2.1 (Zhang et al. 2009). The WRF ARW uses the Arakawa C grid with a terrain-following mass vertical coordinate. For this test, no ocean model was employed, and the ARW was configured using the Lambert-Conformal projection with two, three or four moving, two-way interactive nested domains, as described in Tables 3.10-3.13. All runs used 35 full vertical levels with a model top at 10 hPa.

Table 3.10: PSU configuration used for 13.5 km runs

	DOMAIN 1	DOMAIN 2 (PSU1)
Grid movement	Static	Moving
Horizontal grid spacing	40.5 km	13.5 km
Horizontal dimensions	160 x 121	160 x 121

Table 3.11: PSU configuration used for 4.5 km runs

	DOMAIN 1	DOMAIN 2 (PSU5)	DOMAIN 3 (PSU2)
Grid movement	Static	Moving	Moving
Horizontal grid spacing	40.5 km	13.5 km	4.5 km
Horizontal dimensions	160 x 121	160 x 121	253 x 253

Table 3.12: PSU configuration used for 1.5 km runs

	DOMAIN 1	DOMAIN 2 (PSU4)	DOMAIN 3	DOMAIN 4 (PSU3)
Grid movement	Static	Moving	Moving	Moving
Horizontal grid spacing	40.5 km	13.5 km	4.5 km	1.5 km
Horizontal dimensions	160 x 121	160 x 121	253 x 253	253 x 253

Table 3.13: Physical parameterizations used by PSU

PARAMETERIZATION	SCHEME USED
Cumulus	Grell-Devenyi
Microphysics	WSM6
Radiation	RRTM (longwave) / Dudhia (shortwave)
Planetary Boundary Layer	YSU
Surface Layer	Monin-Obukov
Land Surface	Thermal diffusion

The innermost nest for both the intermediate- and high-resolution configurations was too small to evaluate wind radii, so domains 2 and 3 for the intermediate-resolution configuration and domains 2 and 4 for the high-resolution configuration were submitted to complete the evaluation, with domain 2 only being used to evaluate wind radii.

The initial and boundary condition fields were created by ingesting the GFS analysis from a 12- or 24-h previous case into the WRF Pre-processing System (WPS v2). WRF-Var was used to perturb the initial and boundary conditions to generate a 30-member ensemble. The ensemble was then integrated for several hours until the time when airborne radar observations became available. Super-observations by the NOAA P3 airborne radar were assimilated in each ensemble member while available, after which time the model started a free forecast using the mean of the ensemble as initial conditions and the six-hourly GFS forecasts on a 1-deg grid as boundary conditions. The start of the free forecast occurred between 1 and 6 h before the date specified for the HRH cases. Therefore, a PSU forecast for lead time H actually corresponds to a 1 to 6 h greater lead time.

University of Rhode Island

The University of Rhode Island (URI) group ran the GFDL/URI model (Bender et al. 2007), a primitive equation coupled atmosphere-ocean model formulated in latitude, longitude and sigma coordinates. The atmospheric model domain consists of a triply nested grid configuration, in which the two inner grids are moveable and two-way interactive. The basic properties of the configurations used for the HRH Test are summarized in Tables 3.14-3.16. Both configurations were run with 42 vertical levels and a model top at approximately 0.27 hPa.

Table 3.14: URI configuration used for 1/12 deg grid

	DOMAIN 1	DOMAIN 2	DOMAIN 3
Grid movement	Static	Moving	Moving
Horizontal grid spacing	1/2 deg	1/6 deg	1/12 deg
Horizontal dimensions	151 x 151	67 x 67	61 x 61

Table 3.15: URI configuration used for 1/18 deg grid

	DOMAIN 1	DOMAIN 2	DOMAIN 3
Grid movement	Static	Moving	Moving
Horizontal grid spacing	1/2 deg	1/6 deg	1/18 deg
Horizontal dimensions	151 x 151	67 x 67	91 x 91

Table 3.16: Physical parameterizations used in URI model

PARAMETERIZATION	SCHEME USED
Cumulus	Simplified Arakawa-Schubert
Microphysics	Ferrier
Radiation	Schwarzkopf and Fels (1991) (longwave) / Lacis and Hansen (1974) (shortwave)
Planetary Boundary Layer	Troen and Mahrt (1986)
Surface Layer	Monin-Obukov
Land Surface	Tuleya (1994)

Rather than delivering companion grids for the nested domains that did not cover 15 x 15 degrees, URI filled the portions of the delivered grids that extended beyond the innermost native grids with the appropriate outer domain data interpolated to the 1/12 or 1/18 deg resolution, respectively.

The GFS global analysis and the storm message provided by NHC were used to generate initial conditions for the atmospheric model. An axisymmetric version of the prediction model was used to create an axisymmetric vortex based on the initial storm structure that was estimated from the data in the storm message. The initial conditions were calculated by adding back the model simulated vortex to the environmental fields that were determined from the GFS analysis. Six-hourly GFS forecasts output on 1/2 deg grid were used for lateral boundary conditions.

The Princeton Ocean Model (POM – Yablonsky and Ginis 2008) was run with a 1/6 deg horizontal grid spacing and 23 vertical sigma levels. The ocean model was initialized by a diagnostic and prognostic spinup of the ocean circulations using available climatological ocean data in combination with real-time SST and sea surface height data. During the ocean spinup, realistic representations of the structure and positions of the Loop Currents, Gulf Stream and warm- and cold-core eddies were incorporated.

During one ocean model time step, the atmospheric model is integrated with its own time steps, and the SST is kept constant. The computed wind stress, heat, moisture and radiative fluxes are passed to the ocean model, which is then integrated one timestep to calculate a new SST. The new SST is then used in the following time step of the atmospheric model.

University of Wisconsin-Madison

The UWM group ran the University of Wisconsin Nonhydrostatic Modeling System (UW-NMS). The UW-NWS is a nonhydrostatic mesoscale model built to achieve accuracy simulating the scale-interaction process, primarily through the imposition of enstrophy and kinetic energy conservation. The model is calculated on the Arakawa C grid using a rotated latitude-longitude projection and geopotential height coordinates. For this test, the model was configured with one static domain and either one or two moving, two-way interactive nested domains. The basic properties of these three model configurations are summarized in Tables 3.17-3.20. All three configurations were run with 46 vertical levels and a model top at 22.8 km.

Table 3.17: UWM configuration used for 12 km runs

	DOMAIN 1 (UWM1)
Grid movement	Static
Horizontal grid spacing	12 km
Horizontal dimensions	480 x 430

Table 3.18: UWM Configuration used for 3 km runs

	DOMAIN 1 (UWM5)	DOMAIN 2 (UWM2)
Grid movement	Static	Moving
Horizontal grid spacing	12 km	3 km
Horizontal dimensions	480 x 430	302 x 302

Table 3.19: UWM Configuration used for 1 km runs

	DOMAIN 1 (UWM4)	DOMAIN 2	DOMAIN 3 (UWM3)
Grid movement	Static	Moving	Moving
Horizontal grid spacing	12 km	3 km	1 km
Horizontal dimensions	480 x 430	302 x 302	353 by 353

Table 3.20: Physical parameterizations used in UWM configurations

PARAMETERIZATION	SCHEME USED
Cumulus	None
Microphysics	Flatau (1989) and Tripoli ; 2-moment prognostic scheme (specific humidity and number concentration) for all species except cloud water
Radiation	RRTM (Mlawer et al. 1997; Mlawer and Clough 1997)
Planetary Boundary Layer	K theory (horizontal), TKE (vertical)
Surface Layer	Louis (1979)
Land Surface	1-D soil model (Tremback and Kessler 1985)

Since the innermost nests for the intermediate- and high-resolution configurations were too small to evaluate wind radii, domain 1 for each of these configurations was submitted to complete the

evaluation. Domain 1 from these higher resolution configurations was only used to evaluate the wind radii errors.

Atmospheric fields for the retrospective forecasts were initialized with fields from the GFDL model, whereas lateral boundary conditions were obtained from the GFS forecast output on a 1-deg grid at 6-hr intervals.

The UW-NMS configurations used a 1.5-layer ocean model (mixed-layer, thermocline) forced by 10-m winds that employs a dynamic instability parameterization of entrainment velocity to represent the ocean. The SST, mixed-layer depth, and 20ºC isotherm depth was initialized from HYCOM North and Equatorial Atlantic Prediction System analyses.

4 Test Cases

4.1 Data Inventory

A list of the selected retrospective cases, as well as an inventory of the forecasts each modeling group delivered, are shown in Appendix B. The tracks and intensities of the ten selected storms are shown in Fig. 4.1. Three modeling groups delivered forecasts for all 69 cases, one group delivered forecasts for most of the cases, and two groups were able to deliver 9-17 cases.

4.2 Brief Description of Selected Tropical Cyclones

Wilma 2005

Wilma was a large-envelope cyclone that had a slow initial development over the northwestern Caribbean Sea, followed by the greatest intensification rate ever seen in the Atlantic basin. The small central core evolved into a very large central core as the hurricane crossed the Yucatan Peninsula of Mexico and southern Florida. Wilma became extra-tropical over the western Atlantic. Five forecast periods for this system met the DeMaria-Kaplan RI criteria. Wilma represents perhaps the ultimate in tropical cyclone structure, with the eye contracting to 2-3 n mi in diameter at the time of peak intensity. The system then underwent a series of eyewall replacement cycles to become 50-60 n mi in diameter during its crossing of south Florida. Wilma is an excellent test case for RI, eyewall replacement, land interaction, and extra-tropical transition.

Philippe 2005

Philippe reached hurricane strength east of the Lesser Antilles, and then weakened due to the impact of shear and interaction with an upper-level low. This storm is an excellent test case for arrested development due to shear, which is good for testing false alarms. It should be noted that the GFDL model produced false alarms for RI in the real-time runs.

Rita 2005

Rita developed rapidly over the Florida Straits and the eastern Gulf of Mexico. This development was followed by weakening due to a series of eyewall replacement cycles and increasing vertical shear. Rita made landfall over western Louisiana and eastern Texas. Six forecast periods for this system met the DeMaria-Kaplan RI criteria. Rita is an excellent test case for RI and eyewall replacement. Data collected from the Hurricane Rainband and Intensity Change Experiment (RAINEX) experiment would allow a detailed verification of the model depiction of the cyclone structure.

Karen 2007

Karen was a large-envelope cyclone that formed over the eastern Atlantic Ocean. The system started to develop rapidly and reached hurricane strength just before encountering strong shear, then dissipated over water due to shear. Two forecast periods for this system met the DeMaria-Kaplan RI criteria. Karen represents a case where RI might have been forecast in NWP models, only to be abruptly halted by the increased shear.

Katrina 2005

Katrina formed over the Bahamas, becoming a hurricane just before landfall in southern Florida. This system then intensified over the Gulf of Mexico and underwent an eyewall replacement before reaching category 5 intensity. Katrina then weakened due to another eyewall replacement, and possibly shear and ocean effects before landfall along the northern Gulf coast. Five forecast periods for this storm met the DeMaria-Kaplan RI criteria. Katrina is a good test case for RI, which for this storm was interrupted by an eyewall replacement cycle. It is also a good case for improved track forecasts (particularly before the Florida landfall), land interaction, and oceanic effects (the Loop Current).

Humberto 2007

Humberto developed rapidly before landfall on the Texas coast, going from a tropical depression to a hurricane in less than 24 hours. Two forecast periods for this system met the DeMaria-Kaplan RI criteria. RI started unusually early in the development of this cyclone, making this storm a good candidate for testing how models will forecast RI in weak systems.

Felix 2007

Felix was a small cyclone that moved westward through the Caribbean Sea and strengthened from a tropical storm to a Category 5 hurricane in 48 hours. The hurricane made landfall in Nicaragua and dissipated over Central America. Eight forecast periods for this system met the DeMaria-Kaplan RI criteria. Felix is an excellent test case for RI, as well as a cyclone that underwent a notable eyewall replacement cycle.

Ingrid 2007

Ingrid was a weak system over the tropical Atlantic whose development was limited by shear and dissipated over water. This system is a good test case for model false alarms of RI.

Emily 2005

Emily developed over the tropical Atlantic, became a hurricane near the Windward Islands and then, after an eyewall replacement cycle, became a category 5 hurricane over the Caribbean Sea. Emily made landfall over the Yucatan Peninsula as a major hurricane, weakened, and then re-intensified over the western Gulf of Mexico. At final landfall in Mexico, the hurricane featured a prominent concentric eyewall in both radar and aircraft data. Eleven forecast periods for this system met the DeMaria-Kaplan RI criteria. Emily is a good test case for RI and intensity fluctuations, with 2 eyewall replacement cycles and one over-water weakening phase with no obvious cause. The initial intensification to hurricane strength was accompanied by a position jump, suggesting the vortex re-formed or re-organized.

Ophelia 2005

Ophelia formed from a non-tropical trough and followed a meandering course off the southeastern coast of the U. S. for many days. While there were periods of RI and the peak intensity was only 75 kts, there were numerous fluctuations in intensity. Ophelia eventually passed Cape Hatteras and became extra-tropical as it approached Nova Scotia. This cyclone is the poster child for oceanic effects, with a slow meandering track that brought it across its own cold wake or allowed it to remain stationary long enough to upwell.

5 HRH Evaluation System

The HRH evaluation system (Fig. 5.1) is composed of the following elements: 1) data delivery, 2) GFDL Vortex Tracker, 3) averager script, 4) plotting scripts, 5) display, 6) NHC Verification System, 7) aggregation and statistical significance assessment, 8) RI/RW verification, 9) consistency verification, and 10) archival. The first six processes in this system were run in an automated mode on the NOAA ESRL Linux cluster using the NOAA ESRL Workflow Manager (Harrop et al. 2007), while archival was run as a separate automated task. The initial automated processes were triggered by the delivery of the forecast files, and were run incrementally as the files arrived. The aggregation and statistical significance assessment, RI/RW verification and consistency verification were run after all forecasts were delivered.

1) Data Delivery

The input to the HRH evaluation system consists of gridded data files provided by the modeling groups and Best Track and tcvitals files provided by the NHC. Each gridded data file delivered to the DTC contains the required input fields for the GFDL Vortex Tracker (zonal and meridional wind components and geopotential height at 850, 700 and 500 hPa, zonal and meridional wind components at 10-m, absolute vorticity at 850 and 700 hPa, and mean sea level pressure - MSLP), plus temperature and dewpoint temperature at 2-m, 850, 700 and 500 hPa and optional 1-h accumulated precipitation used for plotting and display. These files are in GRIB1 format and contain forecasts every 30 minutes.

2) GFDL Vortex Tracker

A revised version of the GFDL Vortex Tracker (Marchok 2002) was used to locate the storm in each forecast. GFDL implemented several modifications to make the standard version of the Vortex Tracker suitable for the HRH Test:

- The output format was modified from the standard Automated Tropical Cyclone Forecasting (ATCF) format to allow the forecast lead time field to be expressed in decimal format, making it possible to represent sub-hourly forecast lead times.
- The code was expanded to include the ability to process moving nests.
- Parameters of the Barnes analysis were altered to accommodate high-resolution forecasts.
- Ability to deal with regions of missing data was increased.

In spite of all the enhancements, the tracker was not able to follow all storms in the forecasts processed for the HRH Test. Forecasts for which the storm is weak or disorganized, even if

matching the observed storm, cannot be tracked. If the tracker cannot find the storm at a given forecast lead time, it is not able to locate the storm at longer lead times, regardless of whether the storm becomes organized at longer lead times. Hence, an organized storm at longer lead times may not be tracked, and therefore may not be part of the evaluation. Inventories in Appendix C indicate which forecasts were actually tracked and evaluated.

3) *Averager Script*

Observed and forecasted surface winds for tropical storms can vary significantly over a small time period. The maximum surface wind speed (MSW) contained in NHC Best Track files represents a maximum 1-minute sustained surface winds (MSSW). Due to time constraints for completing this test, it was not practical to require the modeling groups produce 1-minute sustained surface winds. Hence, the verification of MSSW was based on the average of the MSW over a two-hour window centered at the verification time. The 2-hour average was computed using gridded data at 30-minute intervals (i.e., average over five data points). Data at minus (plus) 30 and 60 minutes were not available for forecast lead times at the beginning (end) of the forecast, so a one-sided 1-hour average was computed for these lead times. Only MSWs were averaged; that is, storm location and extent of wind radii were not averaged. The *averager* script in the HRH evaluation system ingests the GFDL Vortex Tracker output in modified ATCF format and outputs averaged files in standard ATCF format, with information restricted to a forecast frequency of six hours.

4) *Plotting Scripts*

Plots for the forecasts submitted by the modeling groups were generated using the NCAR Command Language (NCL). Contour plots for surface fields included: 10-m winds, 2-m temperature and dewpoint temperature, and MSLP. Contour plots for upper air fields included: 850, 700, and 500 hPa temperature, dewpoint temperature and winds. Additionally, plots of track, minimum MSLP, MSW, and the change in MSW over a 24 hour period with respect to lead time were created for all resolutions of a given model. All multi-resolution plots also incorporated the relevant NHC Best Track information.

5) *Display*

All plots generated by the automated evaluation system were made available to HRH participants through the website http://www.dtcenter.org/plots/hrh_test/graphics. This website provides an interface for requesting the multi-resolution plots, a side-by-side display of contour plots by modeling group, as well as the capability to animate loops of a particular type of contour plot.

6) *NHC Verification System*

The NHC Verification System was used to verify the forecasts against the NHC Best Track. Only the tropical portion of the tracks was verified. Each case was processed separately; that is, the input for each run of the NHC Verification System included a single forecast for one model resolution. All forecasts were run through the NHC verification system twice: once for all forecast tracks and a second time for tracks over water only. For the latter, only situations in which both the observed and forecast storm centers are over the ocean were

considered. Variables verified include: storm location, averaged MSW, and extent of wind radii. Metrics generated include: absolute, cross- and along-track error, MSW error, and wind radii error.

7) Aggregation and Statistical Significance Assessment

Results from the individual runs of the NHC Verification System were aggregated using a script in the R statistical language. Given the distribution of errors and absolute errors at a given lead time, several parameters of the distribution were computed: mean, median, quartiles, and outliers. Confidence intervals (CI) on the median were computed using a parametric method (Chambers et al. 1983, McGill et al. 1978). Only lead times and errors for which the distribution contained at least 11 samples are considered in the statistical significance discussions because the error distribution parameters cannot be computed for sample sizes less than 11. Skill scores for all models at all resolutions were also computed against the Decay-SHIFOR5 (OCD5), but will not be discussed in this report. Plots of these scores will be made available on the HRH website.

A pairwise technique was used to address the question of whether the differences between a medium- or high-resolution model configuration and their low-resolution counterpart are statistically significant (SS). For this technique, the absolute error of a given quantity (for example, track error) for a medium- or high-resolution forecast is subtracted from the same metric for the low-resolution forecast. This subtraction is done separately for each lead time of each case, yielding a distribution of forecast error differences. The parameters of this difference distribution are then computed using the same methodology applied to the error distributions for a single model resolution. The pairwise method enables the identification of subtle differences between two error distributions that may go undetected when the mean absolute error (MAE) or root mean square error (RMSE) of each distribution is computed and the overlap of the CIs for the mean is used to ascertain differences (e.g., Lanzante 2005, Snedecor and Cochran 1980).

A statistically significant (SS) difference between the forecast verification metrics of the multiple resolutions for a given model was noted when it was possible to ascertain with 95% confidence that the median of the pairwise differences was not equal to zero. The median was chosen over the mean for this analysis because it is a robust measure, which is appropriate for this test some error distributions differed from normality and presented outliers.

Boxplots provide a concise format for displaying the various attributes of the error distributions computed for the HRH Test. Fig. 5.2 illustrates the basic properties of the boxplots used for this report. The mean of the distribution is depicted as a star and the median as a bold horizontal bar. The 95% CIs are shown as the *waist* or *notch* of the boxplot.

Fig. 5.3, which displays a series of boxplots as a function of forecast lead time for a pairwise difference illustrates the method used to determine if a difference between forecasts for two resolutions is SS. When the median of the distribution is positive (negative), it can be concluded that the higher- (lower-) resolution configuration has smaller error. However,

only when the 95% CIs do not encompass the zero line can the differences between forecasts for multiple resolutions be declared SS. In this example, the differences are SS at lead times 6, 12, 36, 42 and 66 h, with the low-resolution configuration presenting better results for all these lead times.

Since there was interest in knowing which cases were harder to forecast, the outliers of the error single-model and pairwise distributions were recorded. Outliers are defined as any data point that lies more than 1.5 *IQR (inter-quartile range) lower than the first quartile or 1.5*IQR higher than the third quartile.

8) RI/RW

No existing tools were available for assessing changes in forecast skill for RI and RW events resulting from changes in resolution. The RI/RW parameters considered by the DTC included: frequency of occurrence, timing of onset, and event-based contingency table scores for matched pairs with time relaxation.

To explore the properties of the observed and forecasted RI and RW events, total counts of RI and RW events were compiled for the entire sample partitioned by model configuration and resolution. A comparison of the medium or high- and low-resolution configurations in the context of the observed occurrence frequency provides insight into whether the forecasted events occur more frequently, less frequently or at the same frequency as that observed. Two methods for defining an event were considered: 1) episodes – define any sequence of one or more periods of rapid change to be one event, and 2) individual - define every period of rapid intensity change to be an event. The episode approach does not penalize forecasts that capture the occurrence of the event but not the duration, whereas the individual approach considers both the number and duration of the forecasted events.

For evaluating the skill of predicting the onset of RI and RW events, the onset of an event is defined as the hour in which a single isolated event or the first in a sequence of events occurs. The timing performance was evaluated by preparing cumulative frequency plots for a 48-h window centered on the observed onset. For these cumulative frequency plots, the observed events appear as a step function at time zero corresponding to the total number of observed events. The count of onset occurrences predicted by the full set of runs for each model is incremented in the appropriate time period relative to the observed onset. Forecasted events for which the onset is earlier than observed appear as counts for negative lead times and forecasted events for which the onset later than observed appear as counts for positive lead times. For uniformly perfect model timing, the forecasted events would also appear as a step function at time zero. This manner of presentation provides information on both timing errors and the number of missed events, but ignores false alarms.

The following scores (see definitions below) were computed for each model configuration using event-based contingency tables for exact matches between forecast and observed RI and RW events, as well as matched pairs obtained by considering successively longer time relaxation windows: Proportion Correct (PC), Probability of Detection (POD), Critical Success Index (CSI), and False Alarm Rate (FAR). The matched pairs considered for this report do not include missed forecast events stemming from the forecasted track being

shorter than the observed track (i.e., lead times for which the tracker did not produce a fix are not included in the sample). The search for matched pairs with several time relaxation windows (6, 12, 18, 24, 30, 36 h) was performed such that the total number of forecasted events is maintained (i.e., each event is only used once when searching for a match for a given time window). This search is conducted from shortest to longest lead times for a particular forecast. The shifted forecast sequence obtained through this methodology will only improve or leave the forecast skill unchanged.

Contingency Table:

		Observed	
Forecast		Yes	No
	Yes	Hit (a)	False Alarm (b)
	No	Miss (c)	Correct Rejection (d)

Scores:

PC = (a+d)/(a+b+c+d)

POD = a/(a+c)

CSI = a/(a+b+c)

FAR = b/(a+b)

9) Consistency

A methodology and software for verifying consistency was developed at the DTC. For the purposes of this study, consistency was defined as the variability of forecasted storm center among runs of a given model and resolution initialized at various times and valid at the same time. In other words, consistency refers to the differences in forecast at multiple lead times for the same valid time. Higher consistency, or smaller variability from one initialization time to the next, is a desirable property of a set of forecasts. Consistency results for the low- and high-resolutions of each model were inter-compared to determine if higher resolution led to higher consistency. Location was the only variable considered and intensity was not taken into account.

The consistency assessment requires forecast cases with high-temporal frequency. While it is possible to apply this methodology to cases that are 24-h apart, the results are less relevant, since it is the short-term consistency (up to 24-h) that has highest practical operational applicability. Since it was not possible to have all HRH cases be at 6-h intervals because of the effect that temporal correlation would have had on the effective sample size, a single storm was chosen for the consistency assessment. Hurricane Felix was chosen because both NOAA operational hurricane models (GFDL and Hurricane WRF – HWRF) displayed dramatic run-to-run variability (low consistency). For Felix, four of the six models participating in the HRH Test submitted a series of runs that were initialized at 6-hour intervals for a 30 hour period.

Fig. 5.4 illustrates how consistency for a given valid time was assessed by creating a series of ten differences between storm positions forecast at various lead times. Since the forecast lead times increases with valid time, an increase in value of the differences towards later valid times is expected. Plots of differences as a function of lead time were created contrasting the results for the high- and low-resolution configurations.

No comparison between the forecast and track of the storm was performed in the consistency evaluation since other metrics address forecast accuracy. This leaves open the possibility that a model could have very consistent forecasts that all might be in error.

10) Archival

The following files have been archived at the NCAR Mass Storage System for all models at all resolutions: forecast files in GRIB1 format, output from the GFDL Vortex Tracker, averager script, and NHC Verification System, RI/RW and consistency verification, plots, and source code and scripts for the HRH evaluation system.

6 Results

Due to the nature of the selected cases, the sample size for longer lead times falls short of 69 (i.e., the tropical cyclone phase of the storm did not span five days for all the selected cases). The sample size for longer lead times is also impacted by forecasts for which the vortex tracker was not able to produce a fix for all submitted lead times. The current version of the GFDL Vortex Tracker stops as soon as it reaches a lead time for which it fails to obtain an appropriate fix. Hence, forecasts for which the storm is weak and disorganized at early lead times but spin up a well defined tropical cyclone at longer lead times may not be included in the verification. Sample sizes for longer lead times over water are also impacted by the nature of the selected cases (i.e., storm for some cases was over land for longer lead times).

6.1 AOML

An inventory of the data included in the objective verification for AOML is shown in Appendix C (Table C1). A table summarizing those lead times for which the difference between the high and low resolution was found to be SS is shown in Appendix D (Table D1). Although AOML delivered gridded data for all the selected cases, as well as the requested lead times, the tracker was not able to identify the storm for all provided forecasts. The tracker output for one case was also cropped due to concerns about accuracy of tracker fix when storm is near the boundary of the inner nest. A few cases were removed from the sample due to concerns about the accuracy of the tracker fixes, but the tracker stopped early in the forecast for these cases, so this removal had a minor impact on the sample size.

Track

The median of the track errors, as well as the spread in these errors increases with lead time for both of AOML configurations (see Fig. 6.1.1). The track error for the low-resolution configuration undergoes a larger increase than that for the high-resolution configuration, leading to a SS difference for which the high resolution is more accurate for lead times 30 to 48 h (see Table D1 in Appendix D). These SS differences correspond to a track improvement on the order of 10 nm. The distributions for cross- and along track errors are generally centered on zero, except for the along-track high-resolution distribution at the longest lead times (not shown). Once again, the only SS differences for cross- and along-track errors occur in the 30 to 48 h time frame (see Table D1 in Appendix D). Limiting the verification to forecast times for which the storm is over water had little impact on the relationship between the track error distributions for the two configurations (not shown).

Intensity

The median of the absolute intensity error distributions for the two AOML configurations does not exhibit any strong trends with lead time and the increase in spread with lead time is noticeably less than that for the track errors (see Fig. 6.1.2a). The intensity error distributions for the two AOML configurations show that both resolutions tend to underpredict storm intensity for early lead times with the high resolution exhibiting a smaller underprediction (see Fig. 6.1.2b). Intermediate lead times exhibit less of a tendency to underpredict intensity, transitioning to an overprediction by the high resolution for the longer lead times and once again underprediction by the low resolution. SS differences for intensity errors occur for 0 to 6 h and 24 to 30 h with intensity improvement for the high-resolution configuration on the order of 5 kts (see Table D1 in Appendix D). Limiting the verification to forecast lead times for which the storm is over water had little impact on the error distributions for early leads times, whereas the error distributions for intermediate lead times show a tendency for both configurations to overpredict intensity with the high resolution exhibiting a larger overprediction (not shown). This change in the distribution led to SS differences at intermediate lead times that now favor the low-resolution configuration. Sample size for longer lead times over water is insufficient to make any assessments.

Wind Radii

The 34 kt wind radii error distributions show that the high-resolution configuration has a stronger tendency to be positive for most lead times, whereas the errors for the low-resolution configuration are centered on zero for shorter lead times and then become positive but smaller than that for the high resolution at longer lead times. For the 50 kt radii, the errors associated with the high resolution are still generally positive for most lead times with the low resolution only showing a positive trend for the longest lead times. And finally, for the 64 kt radii, the errors associated with the high resolution are still generally positive, but not for the low resolution. It is interesting to note that the radii for all thresholds and quadrants tend to be too large at the initial time. All SS differences for the wind radii parameters favor the low-resolution configuration over that of the high-resolution configuration (see Table D1 in Appendix D), which appears to stem from the high-resolution configuration producing radii that are too large. Due to the nature of the wind radii, the sample size for over water only tends to be too limited to make any valid assessments.

Rapid Intensification and Weakening

Table 6.1 shows the frequency of occurrence totals for RI and RW compiled using both the event and episode methodologies. Going to higher resolution improved the agreement between the frequency of observed and forecasted RI events, whereas it produced only a slight improvement when considering RW events. The episode methodology reveals that the improvement in frequency of occurrence for RI events is partially due to the high-resolution configuration producing more RI episodes then was actually observed. The observed RI and RW counts undergo a larger decrease then the forecasted counts for both resolutions when going from events to episodes, which points to the model is having difficulty sustaining periods of RI and RW regardless of the resolution. The cumulative frequency plot shown in Fig. 6.1.3 indicates the two resolutions have similar timing error distributions with the higher resolution forecasts capturing more of the RI episodes. For this sample of episodes, the forecasted episodes tend to lag the observed RI episodes.

Table 6.1: Total RI and RW counts for events and episodes found in Best Track and the high- and low-resolution AOML forecasts.

	Observed	High Resolution	Low Resolution
RI Events	79	70	30
RI Episodes	27	32	15
RW Events	26	7	2
RW Episodes	18	6	2

The RI event-based scores for the various time relaxation windows are shown in Fig. 6.1.4. Scores for both resolutions generally improve as the time relaxation window is expanded. These scores indicate the high-resolution forecasts produce higher POD and CSI for all time relaxation windows, whereas the PC is better for the low-resolution forecast when considering exact and small time relaxation windows. For larger time relaxation windows, the PC is basically the same for the two resolutions. The relationship between the FAR for the two resolutions varies with time relaxation window. Hence, it is difficult to make any conclusions based on the RI scores as to which resolution produced the most skillful forecast. The number of RW events included in this sample is basically too small to justify considering the scores for this type of event.

Consistency

Subjective inspection of the tracks (Fig. 6.1.5) reveals that the high-resolution configuration have greater consistency owing particularly to one rather errant lower-resolution run.

A plot of the 10 differences that make up the set of distance measurements of consistency for several valid times is shown in Fig. 6.1.6 indicating that consistency is larger for the high-resolution configuration, since the differences in storm location are large at later valid times for the low-resolution configuration. A detailed analysis of the results (not shown) indicated that the large differences seen in Fig. AOM2 for the low-resolution configuration are those involving the run initialized at 06 UTC on 2 September, the deviant run depicted in Fig. AOM1. When this run is excluded, the difference in consistency between the high- and low-resolution configurations decreases, and for some valid times is actually better for the low-resolution run.

Overall Evaluation

Increasing the horizontal resolution of the AOML model configuration reduced track and intensity errors for the short to intermediate lead times and improved the frequency of forecasted RI events, but the consistent degradation of the wind radii errors associated with going to higher resolution, as well as frequency of RI episodes exceeding the observed frequency and the difficulty of sustaining RI and RW periods suggests the improvements in track and intensity may not stem from the higher resolution forecasts producing a more accurate representation of the actual storm structure and evolution.

6.2 MMM

An inventory of the data included in the objective verification for MMM is shown in Appendix C (Table C2). A table summarizing those lead times for which the difference between the high and low resolution was found to be SS is shown in Appendix D (Table D2). Although MMM delivered gridded data for all the selected cases, as well as most of the requested lead times, the

tracker was not able to identify the storm for all provided forecasts. The tracker output for a number of the cases was also cropped due to concerns about accuracy of tracker fix when the storm is near the boundary of the domain or problems with a subset of the delivered fields that may have impacted the accuracy of the tracker fix. A few cases were removed from the sample due to concerns about the accuracy of the tracker fix or grid navigation problems that lead to the tracker failing to identify the storm at the initial time of the forecast.

Track

The median of the track errors, as well as the spread in these errors increases with lead time for both of MMM configurations (see Fig. 6.2.1). The track error for the low-resolution configuration undergoes a larger increase than that for the high-resolution configuration at longer lead times, leading to SS differences for which the high resolution is more accurate for lead times 84 to 114 h (see Table D2). These SS differences correspond to a track improvement on the order of 25 nm. The distributions for cross- and along track errors are generally centered on zero, indicating that either sign error is equally likely (not shown). SS differences for cross- and along-track errors that favor the high-resolution configuration also occur in the 84 to 114 h time frame (see Table D2), whereas one SS difference favoring the low-resolution configuration for along-track errors occurs at 30 h. Limiting the verification to forecast times for which the storm is over water leads to more SS differences favoring the low-resolution configuration for earlier lead times (not shown). These SS differences that favor the low resolution correspond to track improvements that are less than 10 nm. It is interesting to note that the cross- and along-track error distributions for the over water only verification exhibit positive trends for both resolutions for 12-36 h and 36-60 h, respectively (not shown). The sample size for longer lead times is too small to make any determination of the impact of resolution for longer lead times over water.

Intensity

The medians of the absolute intensity error distributions for the two MMM configurations do not exhibit any strong trends with lead time and the spread exhibits only a small increase (see Fig. 6.2.2a). The intensity error distributions show that the low resolution tends to underpredict storm intensity for early lead times and the high resolution tends to overpredict the intensity (see Fig. 6.2.2b). Both resolutions tend to underpredict intensity for longer lead times. Only one SS difference occurs for intensity errors at 18 h with intensity improvement for the low-resolution configuration of less than 5 kts (see Table D2). Limiting the verification to forecast lead times for which the storm is over water once again results in more SS differences at shorter lead times that favor the low-resolution configuration. These SS differences correspond to intensity improvements of less than 10 kts. Sample size for longer lead times over water is insufficient to make any assessments.

Rapid Intensification and Weakening

Table 6.2 shows the frequency of occurrence totals for RI and RW compiled using both the event and episode methodologies. Going to higher resolution improved the agreement between the frequency of observed and forecasted RI events but produced a negligible improvement when considering RW events. The episode methodology reveals that the improvement in frequency of occurrence for RI events is partially due to the high-resolution configuration producing more RI

episodes then was actually observed. The observed RI count undergoes a larger decrease then the forecasted counts for both resolutions when going from events to episodes, which points to the model is having difficulty sustaining periods of RI regardless of the resolution. The cumulative frequency plot shown in Fig. 6.2.3 indicates the two resolutions have similar timing error distributions with the higher resolution forecasts capturing more of the RI episodes. For this sample, the forecasted RI episodes appear to be equally likely to lead or lag the observed episodes.

Table 6.2: Total RI and RW counts for events and episodes found in Best Track and the high-and low-resolution MMM forecasts.

	Observed	High Resolution	Low Resolution
RI Events	77	48	24
RI Episodes	26	27	13
RW Events	24	9	8
RW Episodes	17	5	5

The RI event-based scores are shown in Fig. 6.2.4. Scores for both resolutions generally improve as the time relaxation window is expanded. These scores indicate the high-resolution forecast produces higher POD and CSI for all time relaxation windows, whereas the PC is basically the same for both resolutions. The relationship between the FAR for the two resolutions transitions from being higher for the low-resolution forecasts for small time relaxation windows to being higher for the high-resolution forecasts for the larger time relaxation windows. This trend probably stems from the high-resolution configuration's tendency to produce more RI events than observed. The number of RW events included in this sample is once again too small to justify considering the scores for this type of event.

Consistency

Subjective inspection reveals that the tracks are very similar from run to run for both resolutions (Fig. 6.2.5). The run that is most differentiated from the others is the one initialized at 00 UTC on 3 September, since it deviates to the south of the others and is closer to the Best Track.

The plot of the 10 differences that make up the set of distance measurements of consistency for several valid times (Fig. 6.2.6) also indicates that there is little difference between the high- and low- resolution runs, but it should be noted that the higher resolution runs have slightly lower consistency for some valid times.

Overall Evaluation

Increasing the horizontal resolution of the MMM model configuration reduced the track error for longer lead times, whereas this modification did not have a significant impact on the intensity errors. Conversely, higher resolution improved the forecast's ability to capture the observed frequency of RI events, but not necessarily for the right reason given the forecasted episodes exceeded those observed. Hence, higher resolution for this model configuration does not appear to provide the improvements in intensity predictions that HFIP is looking to address.

6.3 NRL

NRL delivered complete forecasts for 63 cases. However, the GFDL Vortex Tracker was not able to identify the storm for all provided forecasts due to problems with the data, or the storm was weak or moving out of the grid. An inventory of the data included in the objective verification for NRL is shown in Appendix C (Table C3). A table summarizing those lead times for which the difference between the high and low resolution was found to be SS is shown in Appendix D (Table D3).

Track

The median and spread of the track errors increase with lead time for both NRL configurations (see Fig. 6.3.1). Median errors start near zero and grow to 300 nm for the 5-day forecast. Neither configuration displayed systematic SS along-track error. The low-resolution configuration also did not display systematic cross-track error, but the high-resolution configuration produced SS positive cross-track error for some lead times between 30 and 66 h (not shown). The track errors for the two configurations increase at differing rates such that SS track error differences exist at lead times 24, 42, 54, and 96 h, for which the high-resolution configuration is more accurate at 24 h and the high-resolution configuration degrades the forecast for the latter three lead times (Fig. 6.3.2). These SS results correspond to a maximum track error difference of 20 nm. The only SS differences for absolute cross- and along-track errors occur in the last two days of the forecast (see Table D3). High resolution is advantageous for the 120-h cross-track forecast, while it degrades the cross-track forecast for a number of lead times between 72 and 114 h. Limiting the verification to forecast times for which the storm is over water leads to no cases for which the low-resolution configuration produced a better track forecast than the high-resolution configuration. SS differences were found for lead times 24 h (absolute track) and 54 – 60 h (along-track), always favoring the higher-resolution configuration.

Intensity

The absolute intensity errors for the two configurations do not grow with lead time. Rather, their median peaks at the three-day forecast and decreases thereafter. The absolute intensity errors are mainly due to bias: both configurations systematically underpredict intensity out to 90 h, and the low-resolution configuration extends this underprediction to the 5-day forecast (Fig. 6.3.3). SS differences for absolute intensity errors occur at 0-, 6-, 24, and 48-h lead times, with the high resolution improving intensity by up to 5 kts by reducing the underprediction (Table D3). Limiting the verification to forecast lead times for which the storm is over water had little impact on the error distributions for early leads time, while at intermediate lead times, the underprediction was eliminated from the high-resolution configuration. The sample size for longer lead times over water is insufficient to make any assessments.

Wind Radii

The wind radii evaluation indicates that the radii are too large at initialization time for most quadrants and wind thresholds. The radii remain too large for the 34-kt threshold for a number of quadrants and lead times, especially during the first two days of the forecast. Conversely, the 50-and 64-kt radii shift to being too small for a number of lead times and quadrants. SS wind radii error differences occur in about 10% of the lead times/thresholds/quadrants (see Table D3

in Appendix D). All differences except two (54- and 78-h lead times for the 50kt threshold in the SW quadrant) indicate the high resolution degrades the forecast. The sample over water is very limited for the longer lead times but, at the earlier lead times, the overall results are unaltered, with the high resolution leading to degradation in 4 out of the 5 SS results.

Rapid Intensification and Weakening

Table 6.3 shows the frequency of occurrence totals for RI and RW compiled using both the event and episode methodologies. Going to higher resolution improved the agreement between the frequency of observed and forecasted RI events but produced only a small improvement when considering RW events. The number of forecasted RI events and episodes is smaller than observed for both the low- and high-resolution configurations. On the other hand, the difference in the observed and forecast ratio of events to episodes suggests the model has difficulty sustaining periods of RI regardless of the resolution. The cumulative frequency plot shown in Fig. 6.3.4 indicates the two resolutions have similar timing error distributions with the higher resolution forecasts once again being able to capture more of the RI episodes.

Table 6.3: Total RI and RW counts for events and episodes found in Best Track and the high- and low-resolution NRL forecasts.

	Observed	High Resolution	Low Resolution
RI Events	55	17	5
RI Episodes	20	13	4
RW Events	20	11	6
RW Episodes	14	6	5

The RI event-based scores for NRL are shown in Fig. 6.3.5. Scores for both resolutions generally improve as the time relaxation window is expanded to 12 h, with little improvement for larger relaxation times. These scores indicate the high-resolution forecast produces higher POD, CSI and FAR, whereas the PC is basically the same for both resolutions. Hence, these scores do not indicate a clear winner between the two resolutions. The number of RW events included in this sample is once again too small to justify considering the scores for this type of event.

Consistency

The tracks for the two NRL configurations are shown in Fig. 6.3.5. Note that some runs are incomplete (Table D3 in Appendix 3), which considerably reduces the sample size. In both configurations there is considerable spread between the runs initialized at different times and, for a given valid time, runs with later initialization times did in general not come closer to the forecast track.

The plot of the 10 differences that make up the set of distance measurements of consistency for several valid times is shown in Fig. 6.3.6a, indicating that consistency is considerably lower for the high-resolution configuration. However, there are only three complete runs available for the low-resolution configuration, whereas for the high-resolution configuration there are six. Fig. 6.3.6b shows the differences limited to the cases for which both configurations have complete runs. In this analysis, there is not much difference in consistency between the high- and low-resolution configurations.

Overall Evaluation

While the high-resolution NRL configuration had a positive impact on intensity for a few lead times and the number of RI events the model is able to produce, the degradation in track forecasting and wind radii, as well as the increase in FAR, does not support the conclusion that going to higher horizontal resolution will lead to significant improvements in tropical cyclone forecasts.

6.4 PSU

PSU delivered gridded data for 9 cases, all of which had the requested forecast length. The GFDL Vortex Tracker failed to identify the storm for the entire length of the forecast in two cases due to the storm being disorganized or moving out of the grid (see Table C4 in Appendix C for the inventory of data used in the objective evaluation). The number of PSU cases is too small to reach any SS conclusions. Therefore, the results presented in this section should be interpreted with caution. The stratification of the verification to storms over water only further reduces the sample size, so those results will not be presented. Moreover, no consistency verification will be presented for PSU since no Felix cases were run.

Track

The mean track errors for all PSU configurations are near zero at initialization time and grow with lead time, along with the error spread. Mean track errors surpass 400 nm for the low- and intermediate-resolution configurations and 650 nm for the high-resolution configuration (see Fig. 6.4.1). The increase in mean track error is dominated by a single forecast (Wilma initialized 21 October 2005 at 00 UTC) that is particularly bad for the high-resolution configuration, followed by the low-resolution configuration. This single forecast is responsible for the intermediate-(high-)resolution configuration having smaller (larger) mean track errors than the low-resolution configuration. On the other hand, it should be noted that, for longer lead times, the majority of the intermediate-resolution forecasts have larger track errors than the low-resolution forecasts.

The mean along-track errors (not shown) for all resolutions are near zero for the early lead times, transitioning to negative errors at longer lead times, with the largest error growth for the low-resolution configuration. Conversely, the mean cross-track error (not shown) is negative throughout the forecast for all resolutions with the high resolution exhibiting the largest error growth with lead time. The different pace of error growth for each resolution leads to, towards the end of the forecast, an advantage for the intermediate-resolution along-track forecast, and for the low-resolution cross-track forecasts.

Intensity

On average, all resolutions display underprediction in the first three days of forecast and overprediction thereafter (not shown). The mean absolute intensity errors do not grow with lead time (not shown), instead they peak at the 24-h and decrease thereafter. The intermediate- and high-resolutions improve the intensity forecast over the low resolution in the first 2.5 days. Later in the forecast, the spread of the results is too large to reach any conclusions.

Wind Radii

The wind radii error distributions have a very large spread, making the average error difficult to interpret (not shown). The only patterns that may be discerned from the data are that the 34 kt radii tends to be too large in all quadrants and the 50 kt radii in the SE quadrant and the 64 kt radii in the NE quadrant tend to start the forecast with near-zero averages and progress towards radii that are too small. The small sample size and large case–to-case and lead time to lead time variability in the wind radii error differences between the three configurations make it impossible to reach any conclusion about the superiority of one configuration over the other for this parameter.

Rapid Intensification and Weakening

Table 6.4 shows the frequency of occurrence totals for RI and RW compiled by applying both the event and episode methodologies to the PSU cases. Going to higher resolution improved the agreement between the frequency of observed and forecasted RI events, whereas going to higher resolution produced no improvement when considering RW events. The cumulative frequency plot shown in Fig. 6.4.2 indicates all three resolutions have similar timing error distributions for which all onsets tend to be either on time or late, with the higher resolution forecasts being able to progressively capture more of the RI episodes.

Table 6.4: Total RI and RW counts for events and episodes found in Best Track and the high-, intermediate- and low-resolution PSU forecasts.

	Observed	High Resolution	Intermediate Resolution	Low Resolution
RI Events	16	10	6	2
RI Episodes	5	5	3	2
RW Events	10	0	0	0
RW Episodes	4	0	0	0

The RI event-based scores for the PSU configurations are shown in Fig. 6.4.3. Scores for all three resolutions generally improve as the time relaxation window is expanded. These scores indicate increasing horizontal resolution produces progressively higher PC, POD, and CSI, whereas at least for the transition from intermediate to high resolution, the FAR decreases as the resolution increases for exact or short time relaxation windows. Conversely, the FAR increases as horizontal resolution increases when longer time relaxation windows are applied. The number of RW events included in this sample is once again too small to justify considering the scores for this type of event.

Overall Evaluation

The small sample size and large case-to-case and lead time to lead time variability in the performance of the three PSU configurations make it impossible to reach any conclusion about the overall impact of resolution on the forecast skill.

6.5 URI

Although URI delivered gridded data for all the selected cases, the number of hours in the forecast files was incomplete for seven cases. Additionally, the GFDL Vortex Tracker was not able to identify the storm for all provided forecasts. An inventory of the data included in the

objective verification for URI is shown in Appendix C (Table C5). A table summarizing those lead times for which the difference between the high and low resolution was found to be SS is shown in Appendix D (Table D4).

Track

The median track errors for both URI configurations are near zero at initialization time and grow with lead time, along with the error spread, to approximately 200 nm for the 5-day forecast (Fig. 6.5.1). Neither URI configuration present substantial systematic along- or cross-track errors. The absolute track error and the along-track errors are indistinguishable for the two resolutions, but the high-resolution configuration does produce better cross-track forecasts for lead times 48 and 72 h (Table D4). A slightly larger benefit of high-resolution for track forecasting was noted for the verification of tracks over water. The absolute track error had SS lower errors for the high-resolution configuration for lead times 72- and 78-h. While the cross-track error still presents two lead-times for which higher-resolution has less error, the along-track error over water shows advantage of high-resolution over water for the 48-h lead time.

Intensity

The intensity error distributions for the two URI configurations, shown in Fig. 6.5.2, indicate both resolutions tend to underpredict storm intensity for several early lead times (0, 24, and 30 h). Conversely, SS overprediction by both resolutions is noted for several lead times in the fourth day of forecasting (Fig. 6.5.2). The absolute intensity errors do not grow monotonically in time (not shown). Rather, they peak for both resolutions at 3.5 days. Differences in absolute intensity errors between the resolutions (Fig. 6.5.3) are noted at the initialization time (when the higher-resolution minimized the under forecasting) and at the 90- and 96-h lead times (when the higher-resolution exacerbated the overprediction by about 6 kt). Limiting the verification to lead times for which the storm is over water made the intensity errors indistinguishable between configurations for all lead times beyond initialization. However, it should be noted that the sample size for longer lead times over water is insufficient to make any assessments.

Wind Radii

The wind radii error distributions for all thresholds and quadrants show that the radii are SS too small at initialization time and SS too large at most subsequent lead times (not shown). The high-resolution configuration gives SS favorable results at the initialization time by increasing the radii size for the 50 kt threshold in the NE quadrant and for the 34 kt threshold in the SE quadrant (see Table D4 in Appendix D). For longer lead times, about a quarter of the results display SS differences, with all of them favoring the lower-resolution setup. These overall results are upheld when verification is conducted over water, with the caveat that the sample size for longer lead times over water is insufficient to make any assessments.

Rapid Intensification and Weakening

Table 6.5 shows the frequency of occurrence totals for RI and RW compiled by applying both the event and episode methodologies to the URI cases. Going to higher resolution had negligible impact on the frequency of forecasted RI and RW events and episodes. The cumulative frequency plot shown in Fig. 6.5.4 indicates both resolutions have similar timing error

distributions for which timing errors for onset are equally likely to be early or late, with the lower resolution forecasts progressively capturing more of the RI episodes.

Table 6.5: Total RI and RW counts for events and episodes found in Best Track and the high- and low-resolution URI forecasts.

	Observed	High Resolution	Low Resolution
RI Events	94	33	30
RI Episodes	30	22	22
RW Events	27	2	1
RW Episodes	19	2	1

The RI event-based scores for the URI configurations are shown in Fig. 6.5.5. Scores for both resolutions generally improve as the time relaxation window is expanded, with the most notable improvement being connected to decreases in FAR. Score differences between the two URI configurations are rather small, with the most notable difference being associated with FAR for longer time relaxation windows for which the low resolution produces better scores. The number of RW events included in this sample is once again too small to justify considering the scores for this type of event.

Consistency

The tracks for the two URI configurations are shown in Fig. URI1. Subjective evaluation of the track forecasts shows that runs for both resolutions appear to be quite consistent, with the largest deviation occurring for the case initialized at 06 UTC on 2 September. Following this initialization time, the tracks converged toward the observed track for both resolutions, a bit faster for the high-resolution run.

The plot of the 10 differences that make up the set of distance measurements of consistency for several valid times (Fig. URI2) suggests that the two runs are very similar. The greater differences occur for the lower resolution runs at the longer lead times (later valid times), which reflects some difference in behavior of the case initialized at 06 UTC on 2 September. But overall these differences are rather small.

Overall Evaluation

The higher-resolution URI configuration did not substantially improve the errors in track or intensity and it increased the wind radii errors. Increasing horizontal resolution also had a negligible impact on frequency of RI and RW events and episodes and the only notable difference in RI event-based scores favored the low-resolution configuration.

6.6 UWM

UWM delivered two resolutions (9 and 3 km) for 17 of the selected cases and an additional resolution (1 km) for 2 of the selected cases. The GFDL Vortex Tracker was able to identify the storm for all provided forecasts, except the two Emily cases (see Table C6 in Appendix C). In addition, the tracker output for the Emily cases was cropped to the lead times indicated due to concerns about the accuracy of tracker fix as the storm made landfall. An error in the post-

processing prior to delivery to the DTC resulted in an inappropriate assignment of wind values over the higher elevation regions that are used by the tracker to obtain a fix. While the 17 cases for the low- and intermediate-resolutions meets the cut-off used for assessing statistical significance, the number of cases for longer lead times does not meet the cut-off due to the shorter lead times for some of the submitted cases and the tracker stopping early and needing to be cropped for the two Emily cases. Regardless of whether the number of cases for a particular lead time meets the cut-off, the sample size for this modeling group should not be viewed as sufficient to make any strong conclusions. Limiting the verification to forecast lead times over water further exacerbates the sample size dilemma. Hence, the following discussion only addresses verification results that include all forecast tracks. Note that no consistency verification will be presented for UWM since no Felix cases were run.

Track

The median of the track errors for the low- and intermediate-resolution UWM configurations increases with lead time, whereas the spread in these errors does exhibit a consistent increase with time (see Fig. 6.6.1). The spread for the low- and intermediate-resolution UWM configurations cross- and along-track error distributions increase slightly with lead time, but the CIs for these distributions always include zero. Only one SS difference between the low- and intermediate-resolution configurations occurs for track and cross-track error (78 h) and one for along-track (60 h - see Table D5 in Appendix D). All of these SS differences favor the low over the high resolution. The small sample of forecasts for the high resolution configuration also favors the performance of the low-resolution configuration over that of the higher resolution.

Intensity

The intensity error distributions for the low- and intermediate-resolution UWM configurations show that the low resolution tends to underpredict storm intensity for most lead times, whereas the intermediate resolution tends to overpredict the intensity for shorter lead times and then underpredict for longer lead times but to a lesser extent than the low resolution (see Fig. 6.6.2). These trends in error distributions are such that five SS differences favoring the intermediate resolution occur in the 48 to 84 h time frame (see Table D5 in Appendix D). When the differences between the high- and low-resolution errors for the small sample are the same sign, the errors for the low-resolution configuration are smaller.

Wind Radii

All wind radii error distributions with a sufficient sample size show that the low- and intermediate-resolution UWM configurations tend to overpredict the wind radii parameters, except at the initial time for which all radii are too small (not shown). Almost all the SS differences for the wind radii parameters favor the low-resolution configuration over that of the intermediate-resolution configuration (see Table D5 in Appendix D).

Rapid Intensification and Weakening

Table 6.6 shows the frequency of occurrence totals for RI and RW compiled using both the event and episode methodologies for UWM low-and intermediate-resolution configurations. The sample is too small for the high-resolution configuration to include in this assessment. Going to

higher resolution improved the agreement between the frequency of observed and forecasted events for both RI and RW. The episode methodology reveals that the improvement in frequency of occurrence for RI events is partially due to the intermediate-resolution configuration producing more RI episodes then was actually observed. The larger decrease in observed numbers when going from events to episodes also suggests to the model has difficulty sustaining periods of RI and RW regardless of the resolution. The cumulative frequency plot shown in Fig. 6.6.3 indicates the two resolutions have similar timing error distributions, but the higher resolution forecasts are able to capture all of the observed RI episodes, whereas the low-resolution forecasts are only able to capture about half of the observed RI events.

Table 6.6: Total RI and RW counts for events and episodes found in Best Track and the intermediate- and low-resolution UWM forecasts.

	Observed	Intermediate Resolution	Low Resolution
RI Events	46	32	12
RI Episodes	13	18	11
RW Events	18	6	1
RW Episodes	12	5	1

The RI event-based scores for the UWM configurations are shown in Fig. 6.6.4. Scores for both resolutions generally improve as the time relaxation window is expanded, with the time relaxation having a larger impact on high-resolution forecasts. These scores indicate increasing horizontal resolution produces higher POD and CSI, as well as higher FAR. The PC favors low-resolution for exact matches, whereas longer time relaxation windows lead to PC that favors higher resolution. The number of RW events included in this sample is once again too small to justify considering the scores for this type of event.

Overall Evaluation

While the intermediate resolution UWM configuration had a positive impact on intensity for several lead times and the number of RI events the model is able to produce, the degradation in track forecasting and wind radii, as well as the increase in FAR, do not support the conclusion that going to higher horizontal resolution will lead to significant improvements in tropical cyclone forecasts.

7 Discussion

Over the past year the DTC conducted an evaluation of the impacts of high horizontal resolution on tropical cyclone forecasting. This evaluation is based on retrospective forecasts submitted by six modeling groups where each group was required to submit forecasts for at least two resolutions: 1) a baseline configuration with grid spacing comparable to the current operational resolution (~9 km) and 2) a configuration with grid spacing finer than current operational resolution (6 to 1 km). Sixty-nine cases for ten Atlantic storms were analyzed through several objective metrics, including track and intensity errors, RI/RW distributions and scores, and consistency. The overall conclusion, presented in the Executive Summary, is that high horizontal resolution for the model configurations considered in this test did not substantially improve the accuracy of tropical cyclone forecasts. Improvements were noted for some metrics and lead times for every model, but the overwhelming majority of results did not show statistically

significant differences between forecasts at different resolutions with some metrics actually pointing to a degradation of the forecast by going to higher horizontal resolution.

It is important to keep in mind that the models used for this test may simply have been configured in a manner that the potential benefits of high resolution were not able to be realized. These limitations may stem from inadequate physics parameterization and/or initialization techniques. Conversely, the grid spacing necessary to show the benefits of high resolution may not have been reached in this study. Recent studies (e.g., Rotunno et al. 2009) suggest that much higher resolution, on the order of hundreds of meters, might be needed to represent inner core processes important to storm evolution. Additionally, it is possible that higher vertical resolution is required in addition to finer horizontal grid spacing.

We recommend diagnostic studies be conducted for a small sample of cases to determine whether important processes are missing in the forecast. Once those are identified and addressed by the use of alternative physics suites and/or initialization techniques, new comprehensive tests can be conducted for which it is possible the benefits of high-resolution may be realized.

Based on our experience with this data set, it is recommended that a sample larger than 69 cases be used for future tests, taking into consideration that not all forecast initialization times have a corresponding 5-day Best Track due to storm dissipation and only the tropical phase of the storms being verified. Note that the sample is further decreased by the inability of the tracker to produce a fix for the vortex in every forecast.

While GFDL made numerous critical improvements to their vortex tracker to adapt it to high-resolution, issues still remain. One major short-coming is that the tracker is not able to produce fixes for subsequent lead times once it encounters a lead time for which it fails to produce a fix. Due to this limitation, correct forecasts of disorganized storms that progress to stronger storms cannot be verified, potentially introducing a skewness in the forecast sample. In the course of the analysis of the forecasts, it was also found that the minimum native grid size of 3 x 3 degrees requested for this test is not adequate for vortex tracking in many situations, and the minimum grid size of 15 by 15 degrees for extent of radii evaluation is also not enough to capture the size of some large storms. Therefore, it is recommended that future evaluation grids be at least 20 x 20 degrees where the post-processed grids for nested domains that do not meet this minimum size criterion are filled with coarser resolution information where necessary, eliminating the need for the "companion" grids used for radii evaluation in this test.

The DTC tropical cyclone verification system is composed of well-established methods and packages (such as the NHC Verification software) and new exploratory and experimental methods developed by the DTC (such as the RI/RW evaluation). Additionally, the use of the pairwise technique to compare metrics for two configurations when a homogeneous sample (same cases for the two configurations) is available represents a new application of an established statistical technique. Given the non-normal distribution of the errors for various metrics, especially for the extent of wind radii, the use of a robust metric such as the median is preferable over the mean to avoid skewing the results towards a few outliers. To avoid over-emphasizing a few points with large errors, metrics such as ME and RMSE were intentionally left out of this report because they are difficult to interpret. Finally, this evaluation incorporated, whenever possible, a measure of uncertainty to indicate the degree of confidence in the results.

Given the large variability of errors from case to case, and the limited sample size in this test, results must be interpreted cautiously, which can be done when the degree of confidence in the metrics is indicated.

The consistency verification used in this report was useful to give an indication of the differences in run-to-run variability stemming from changes in horizontal resolution. However, the sample size was too small to reach any conclusions, so the consistency results in this report should be seen more as a demonstration of a tool than actual robust results. Since it is known that shorter range consistency is most important to forecasters, we recommend future tests use a larger sample and extend the consistency analysis to contain stratification based on the length of time between runs, so that the consistency between runs that are six, twelve, eighteen etc. hours apart can be depicted. Additionally, a product such as "dProg/dt" can be created in order to evaluate whether the models converge on a solution as the lead time decreases. This product already exists in the NWS Workstation for some fields, and could be extended to feature-base hurricane forecasting.

The tools developed to explore the properties of the forecasted RI and RW events provided interesting insights into the impact of resolution on forecasts of rapid intensity change. Conversely, the application of these tools brought to light short-comings with respect to the test plan design, as well as the information these approaches provide. One major short-coming of the test plan is the selected cases did not provide an adequate sample of RI and RW events for making clear assessments. Given the threshold nature of this metric, it would be useful to have tools that investigate the correlation between the temporal evolution of the observed and forecasted intensity changes in a context that would provide information on whether the forecast totally misses the intensity trend, simply falls slightly short of the threshold criterion, or produces multiple episodes during a single observed episode due to small changes in rate of intensity change when near the threshold. More sophisticated matching and time relaxation methodologies for looking at timing errors might also provide useful information of this type of event.

The HRH test was the first large enterprise of the HFIP program. In addition to assessing of the potential benefits of high-resolution forecasting, this test served as a catalyst for the bringing together a large community of forecasters and scientists of diverse backgrounds (meteorologists, oceanographers, and statisticians, among others) working both for the operational and research and development sector, in a collaboration to improve forecasting. Finally, the test led to the creation of an infrastructure and expertise at the DTC for the evaluation of tropical cyclone forecasts and generated a dataset that is available to the community and can be explored further in comparisons against current operational models, and in studies of resolution, inter-model comparisons, and ensemble forecasting.

Acknowledgments: We would like to thank Tim Marchok for his vital assistance with the GFDL Vortex Tracker, Diane Stokes for her assistance with retrieving archived GFS data, and James Franklin for his assistance with NHC Verification System and various other aspects of tropical cyclone verification.

References

Braun, S. A., M. T. Montgomery, and X. Pu, 2006: High-resolution simulation of Hurricane Bonnie (1998). Part I: The organization of eyewall vertical motion. *J. Atmos. Sci.*, 63, 19–42.

Bender, M.A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane-ocean prediction system and summary of its performance. *Mon. Wea. Rev.*, **135**, 3965-3989.

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey, 1983: *Graphical Methods for Data Analysis.* Wadsworth & Brooks/Cole Publishing Company.

Chen, S. S., J. F. Price, W. Zhao, M. A. Donelan, and E. J. Walsh, 2007: The CBLAST-Hurricane Program and the next-generation fully coupled atmosphere-wave-ocean models for hurricane research and prediction. *Bull. Amer. Meteor. Soc*, **88**, 311-317.

Davis, C., W. Wang, S. Chen, Y. Chen, K. Corbosiero, M. DeMaria, J. Dudhia, G. Holland, J. Klemp, J. Michalakes, H. Reeves, R. Rotunno, and Q. Xiao, 2008: Prediction of landfalling hurricanes with the advanced hurricane WRF model. *Mon Wea. Rev*., **136**, 1990-2005.

Donelan, M. A., B. K. Haus, N. Reul, W. J. Plant, M. Stiassnie, and H. C. Graber, 2004: On the limiting aerodynamic roughness of the ocean in very strong winds. *Geophys. Res. Lett.*, **31,** L18306, doi:10.1029/2004GL019460.

Fairall, C., F. Bradley, D. P. Rogers, J. B. Edson, and G. S. Young, 1996: Bulk parameterization of air–sea fluxes for Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment. *J. Geophys. Res.*, **101,** 3747–3764.

Flatau, P., G. J. Tripoli, J. Verlinde, and W. Cotton, 1989: The CSU RAMS Cloud Microphysics Module: General Theory and Code Documentation. Technical Report 451, Colorado State University, 88 pp. [Dept. of Atmos. Sci., Colo. State Univ., Fort Collins, CO, 80523.]

Harshvardhan, R. Davies, D. Randall, and T. Corsetti, 1987: A fast radiation parameterization for atmospheric circulation models. *J. Geophys. Res.*, **92**, 1009-1015.

Hodur, R.M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.*, **135**, 1414-1430.

Gopalakrishnan, S. G., N. Surgi, R. Tuleya, and Z. Janjic, 2006: NCEP's two-way-interactive-moving-nest NMM-WRF modeling system for hurricane forecasting. Preprints, *27th Conference on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc., Ar. 7A.3.

Harrop C., L. Bernardet, M. Govett, J. Smith, and S. Weygandt, 2007: Workflow Management System for Automating Weather and Climate Simulations. *8th WRF Users' Workshop,* 11-15 June 2007, Boulder, CO. http://www.mmm.ucar.edu/wrf/users/workshops/WS2007/abstracts/p6-1_Harrop.pdf

Hendricks, E. A., M. T. Montgomery, and C. A. Davis, 2004: On the role of vortical hot towers in hurricane formation. *J. Atmos. Sci.*, **61**, 1209–1232.

Janjic, Z. I., J. P. Gerrity Jr., and S. Nickovic, 2001: An alternative approach to nonhydrostatic modeling. *Mon. Wea. Rev.*, **126**, 2599-2620.

Jin, Y., W. T. Thompson, S. Wang, and C.-S. Liou, 2007: A numerical study of the effect of dissipative heating on tropical cyclone intensity. *Wea. Forecasting*. **22,** 950-966.

Lacis, A. A., and J. E. Hansen, 1974: Parameterization for the absorption of solar radiation in the earth's atmosphere, *J. Atmos. Sci.*, **31**, 118-133.

Lanzante, J. R., 2005: A cautionary note on the use of error bars. *J. Climate*, **18**, 3699-3703.

Louis, J.-F., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Bound. Layer. Meteor.*, **17**, 187-202.

Marchok, T. P., 2002: How the NCEP tropical cyclone tracker works. Preprints, *25th Conf. Hurr. Trop. Meteor.*, 29 April – 3 May 2002, San Diego, CA, 21-22.

McGill, R., J. W. Tukey, and W. A. Larsen, 1978: Variations of box plots. *The American Statistician*, **32**, 12–16.

Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono and S.A. Clough, 1997a: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16,663-16,682.

Mlawer, E.J., and S.A. Clough, 1997b: On the extension of rapid radiative transfer model to the shortwave region, in Proceedings of the 6th Atmospheric Radiation Measurement (ARM) Science Team Meeting, U.S. Department of Energy, CONF-9603149.

Powers, J. G., and C. A. Davis, 2002: A cloud-resolving, regional simulation of tropical cyclone formation. *Atmos. Sci. Lett*., doi.10.1006/asle.2002.0054

Rotunno, R., T. Chen, W. Wang, C. Davis, J. Dudhia and G.J. Holland, 2009: Large-eddy Simulation of an Idealized tropical Cyclone. *Bull. Amer. Meteor. Soc.*, (in press).

Rutledge , S. A., and P. V. Hobbs, 1983: The mesoscale and microscale structure of organization of clouds and precipitation in midlatitude cyclones. VIII: A model for the "seeder-feeder" process in warm-frontal rainbands. *J. Atmos. Sci.*, **40**, 1185-1206.

Schwarzkopf, M.D., and S.B. Fels, 1991: The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *J. Geophys. Res.*, **96**, 9075-9096.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3. NCAR Technical Note TN-468+STR. 113 pp

Snedecor, G.W., and W.G. Cochran, 1980: Statistical methods. Iowa State University Press., pp. 99-100.

Trembeck, c. J., and R. Kessler, 1985: A surface temperature and moisture parameterization for use in mesoscale numerical models. Preprints, *Seventh Conf. on Numerical Weather Prediction*, Montreal, Amer. Meteor. Soc.

Tripoli, G. J., 1992: A Nonhydrostatic Mesoscale Model Designed to Simulate Scale Interaction. *Mon. Wea. Rev.*, **120**, 1342-1359.

Troen, I.B. and Mahrt, L., 1986: A Simple Model of the Atmospheric Boundary Layer: Sensivity to Surface Evaporation. *Bound.-Layer Meteor.* **37**, pp. 129-148.

Tuleya, R.E., 1994: Tropical storm development and decay: Sensitivity to surface boundary conditions. *Mon. Wea. Rev.*, **122**, 291-304.

Tuleya, R. E., and Y. Kurihara, 1978: A numerical simulation of the landfall of tropical cyclones. *J. Atmos. Sci.*, **35**, 242-257. 2001: An alternative approach to nonhydrostatic modeling. *Mon. Wea. Rev.*, **126**, 2599-2620

Wang, S., Q. Wang, and J. Doyle, 2002: Some improvement of Louis surface flux parameterization. Preprints, *15th Symp. On Boundary Layers and Turbulence*, Wageningen, Netherlands, Amer. Meteor. Soc., 547–550.

Yablonsky, R. M., and I. Ginis, 2008: Improving the ocean initialization of coupled hurricane-ocean models using feature-based data assimilation. *Mon. Wea. Rev.*, **136**, 2592-2607.

Yau, M. K., Y. Liu, D.-L. Zhang and Y. Chen. 2004: A multiscale numerical study of Hurricane Andrew (1992). Part VI: Small-scale inner-core structures and wind streaks. *Mon. Wea. Rev.*, **132**, 1410–1433.

Yeh, K.-S., X. Zhang, S. Gopalakrishnan, S. Aberson, and R. Rogers, 2009: The AOML/ESRL Hurricane Research System: Performance in the 2008 hurricane season. *Submitted to Journal of Marine Geodesy.*

Zhang, Fuqing, Y. Weng, J. Sippel, Z. Meng, and C Bishop, 2009: Cloud-resolving Hurricane Initialization and Prediction through Assimilation of Doppler Radar Observations with an Ensemble Kalman Filter. *Mon. Wea. Rev.* In Press.

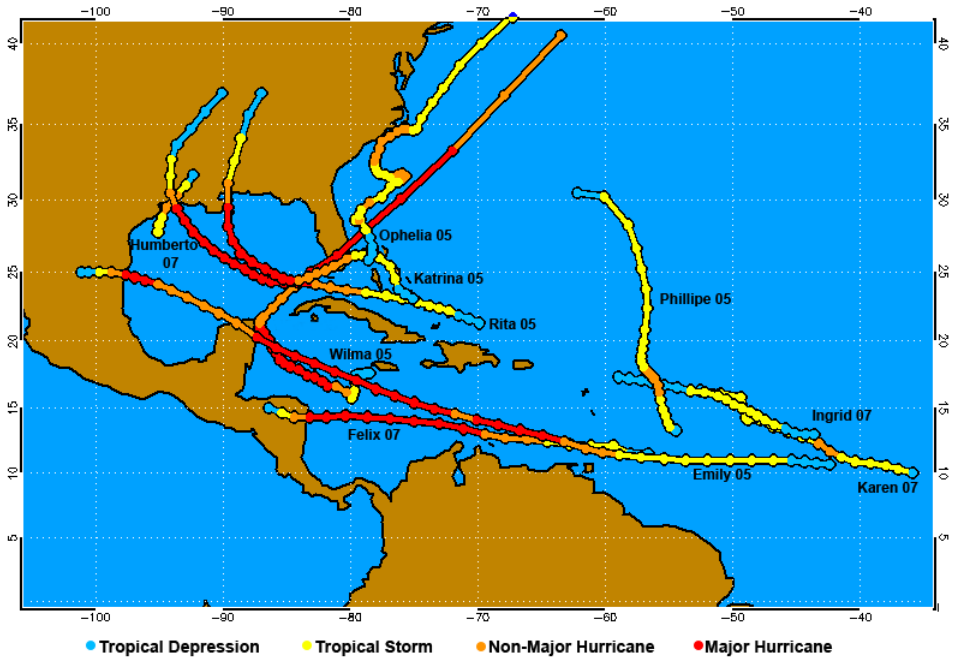


Figure 4.1: Track and intensity of ten storms selected for HRH Test.

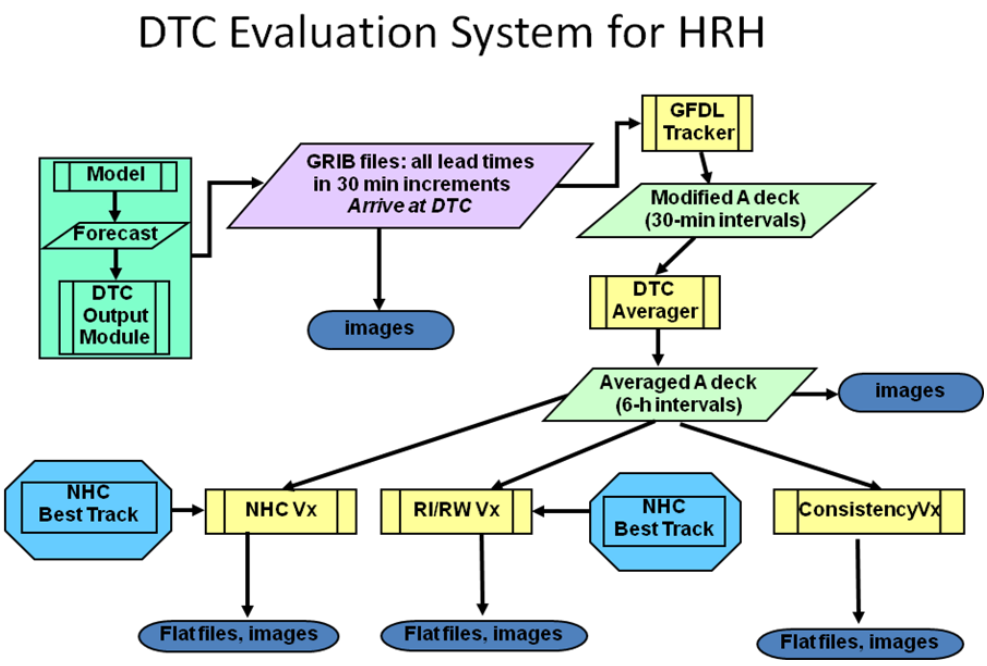


Figure 5.1: Schematic of DTC evaluation system for HRH.

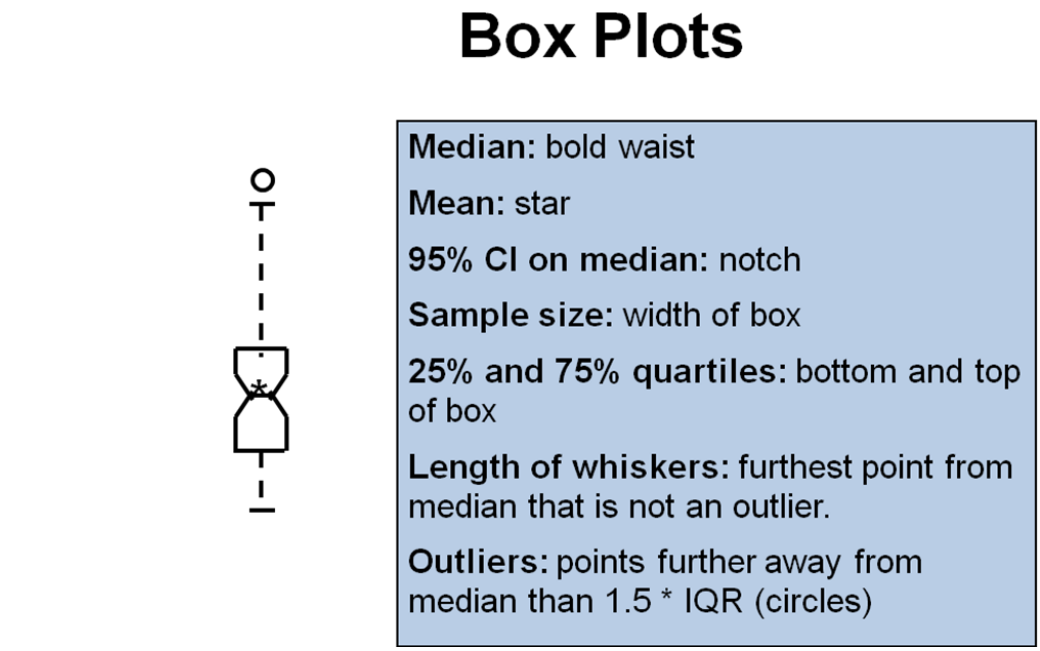


Figure 5.2: Description of the boxplot properties.

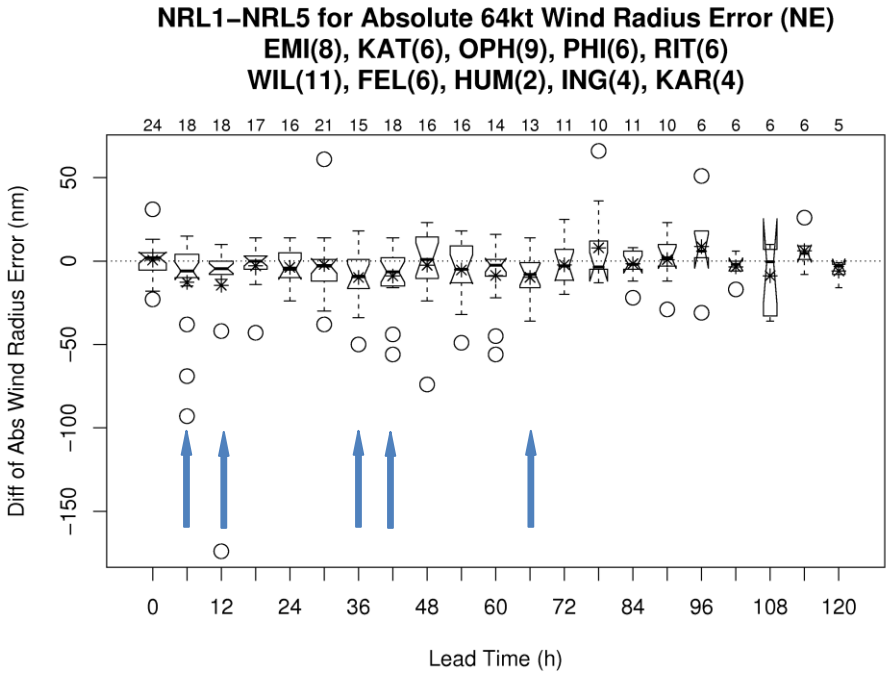


Figure 5.3: Sample pairwise difference boxplot used to identify SS differences. The blue arrows highlight the lead times for which the differences are SS. The numbers above the plot indicate the number of cases in each distribution.

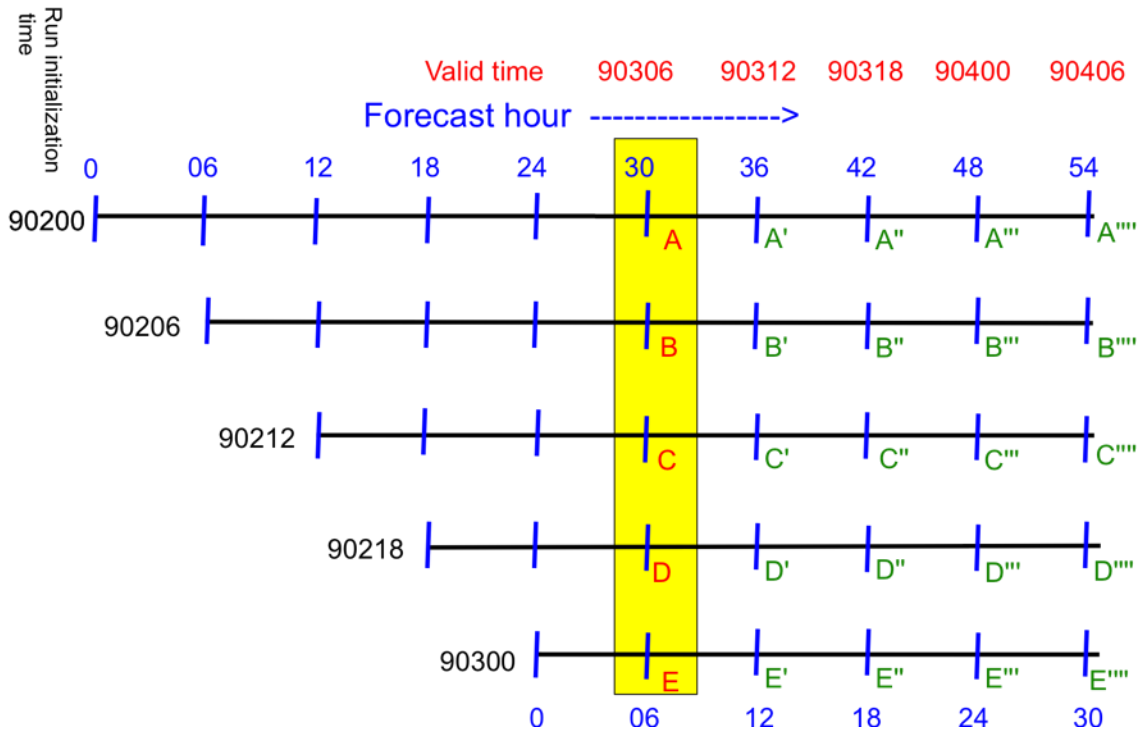


Figure 5.4 Black lines represent the various Hurricane Felix forecasts of a given model resolution, each initialized 6 h after the previous one. Model initialization times are in black on the left of the black lines, following the convention mddhh – month, day, and UTC time of initialization. The valid times (mddhh) for each forecast are in red on the upper part of the figure, while the lead times for the first/last run are depicted in blue above/below the corresponding black line. For each valid time, 10 differences in the position of the center of the storm can be computed, by permutating the runs. For instance, the case highlighted in yellow, valid at 90306, the following differences can be computed between runs initialized at 90200 (A), 90206 (B), 90212 (C), 90218 (D), and 90300 (E): A-B, B-C, C-D, D-E, A-C, A-D, A-E, B-D, B-E, C-D, C-E, D-E.

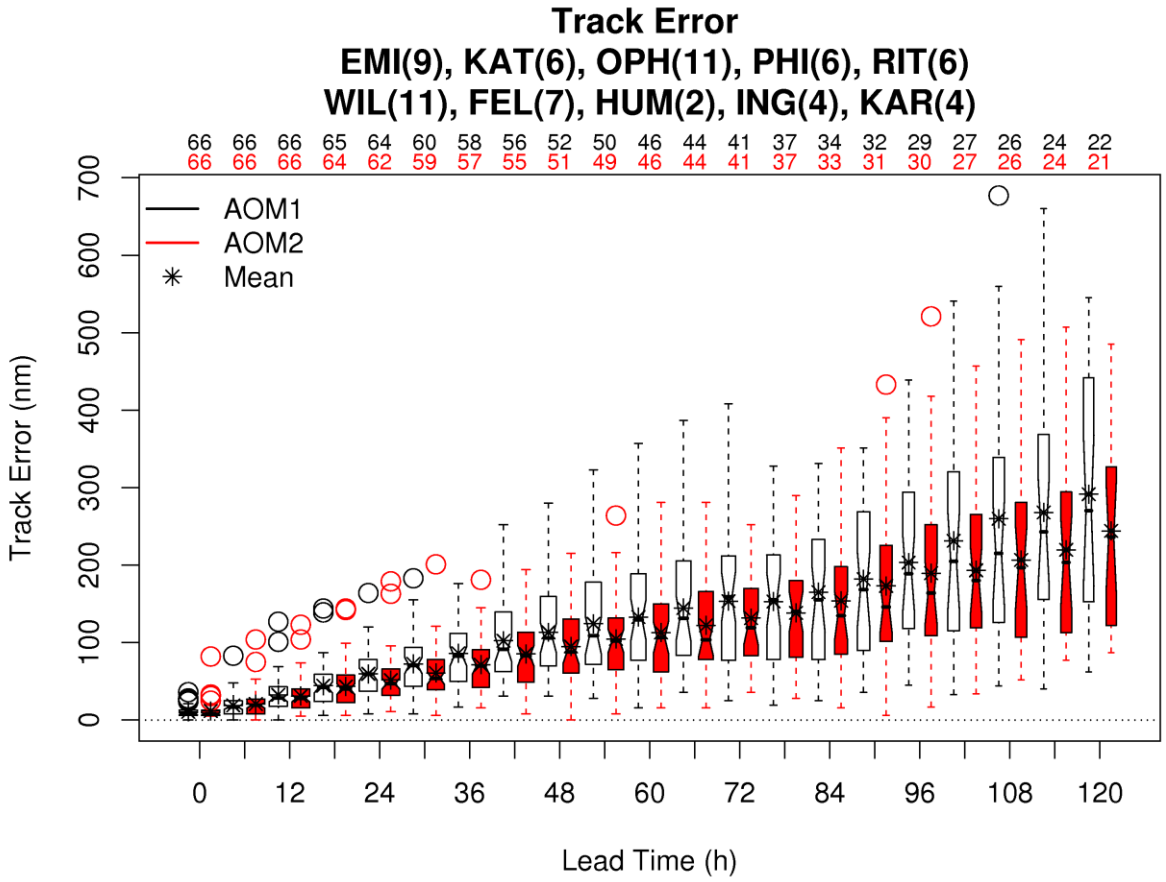
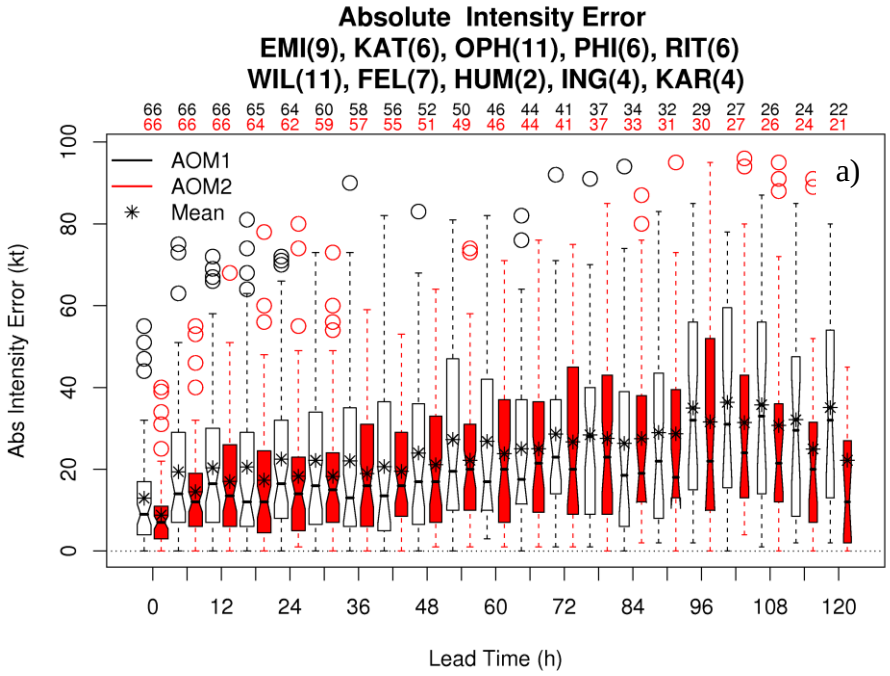


Figure 6.1.1: Track error distributions with respect to lead time for the high- and low-resolution AOML model configurations.



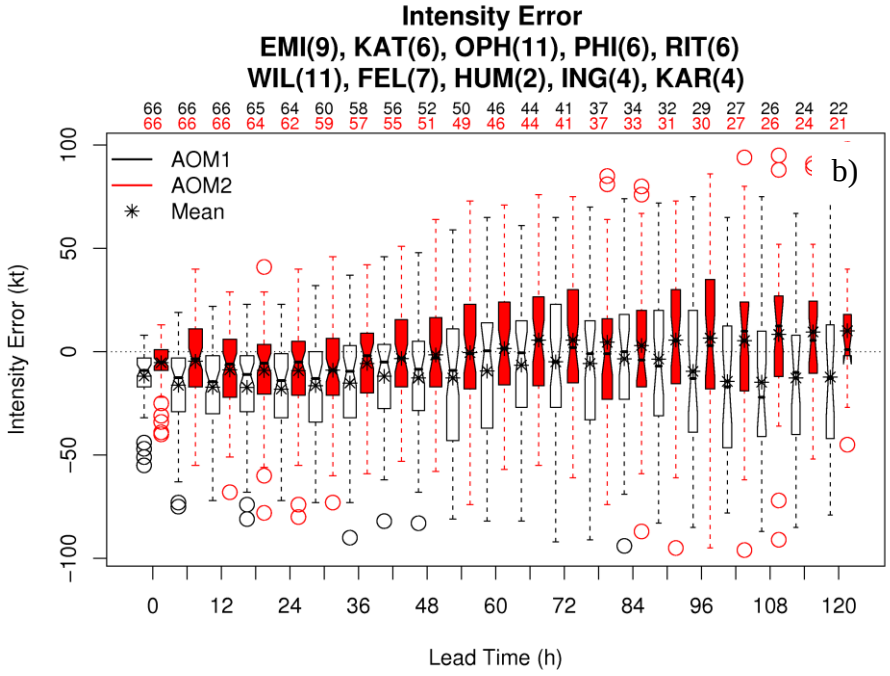


Figure 6.1.2: Same as Fig. 6.1.1, except for a) absolute intensity error and b) intensity error.

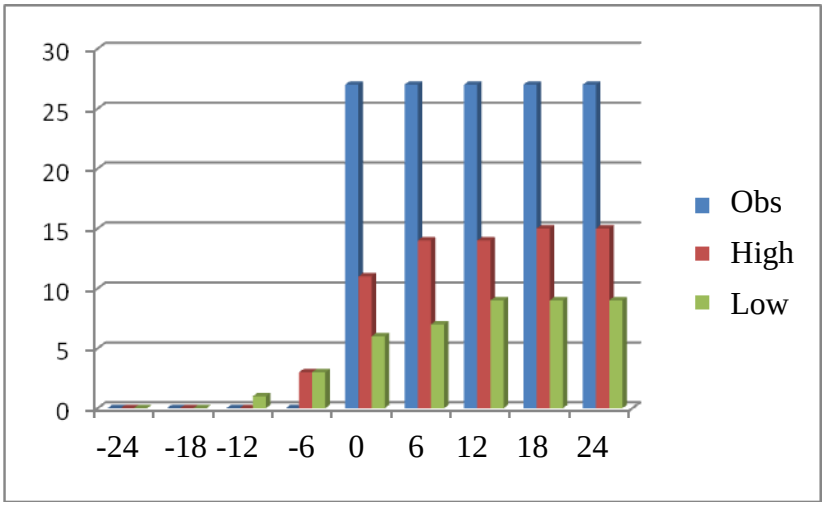


Figure 6.1.3: Cumulative counts of RI events for AOML composited relative to the observed onset times. Blue indicates observations, red are high-resolution events, and green are low-resolution events. Perfect forecasts would be equivalent to the blue bars.

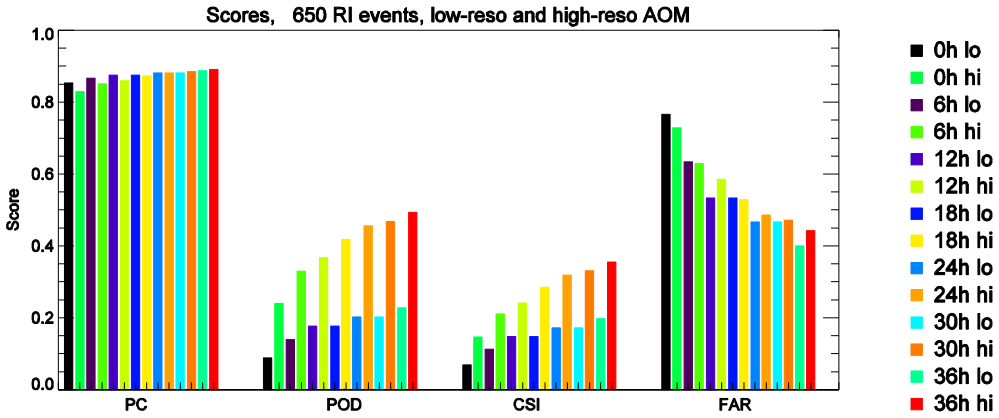


Figure 6.1.4: RI event-based scores for AOML. Columns for each score progress from exact match on the left to increasing time relaxation on the right for column pairs corresponding to low- and high-resolution configurations.

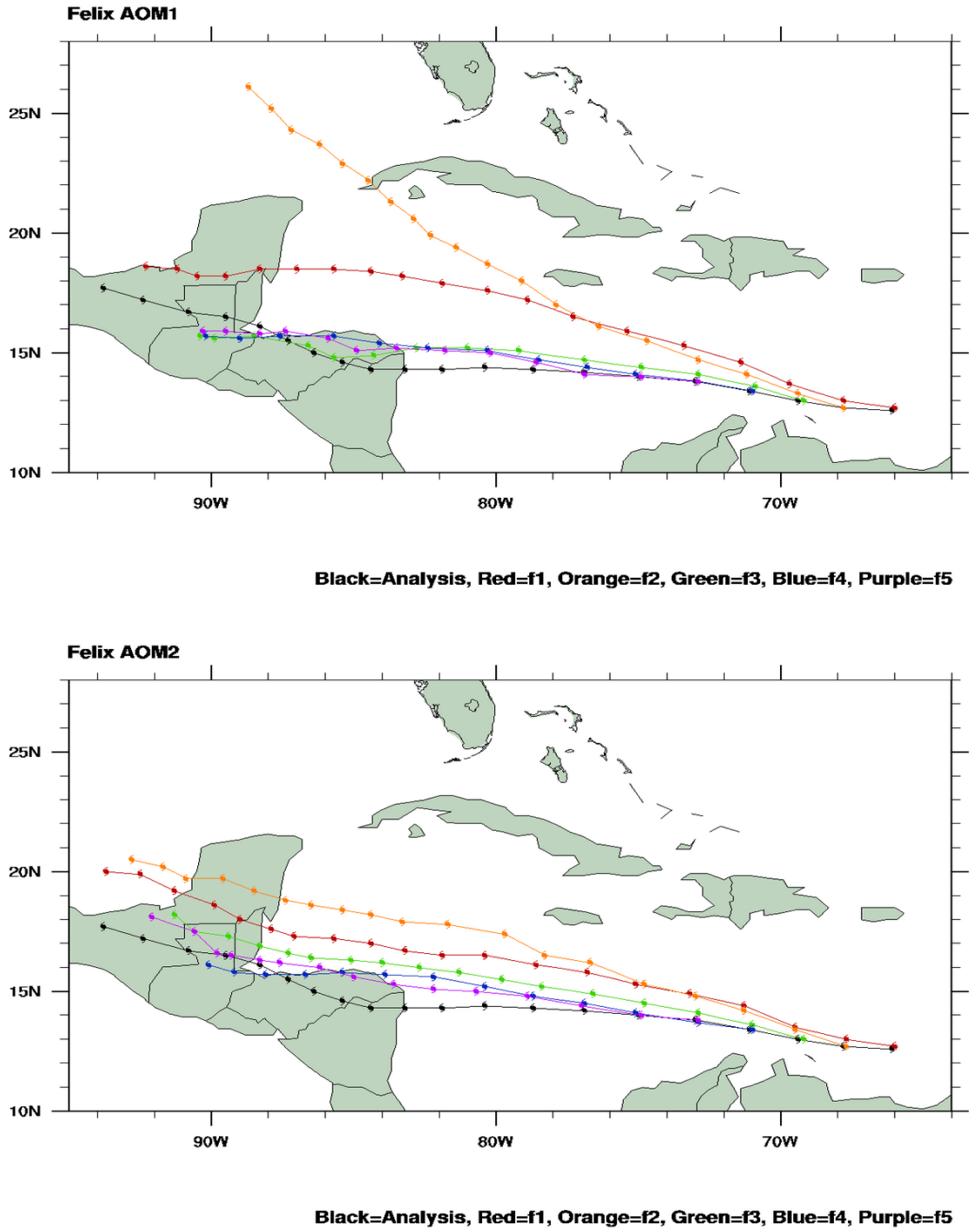


Figure 6.1.5: Tracks for the a) low- and b) high-resolution AOML forecasts for Felix. Red is forecast iniotialized at 00 UTC on 09/02/07, and orange, green, blue and purple correspond to subsequent initializations at 6-h intervals. Black like is Best Track.

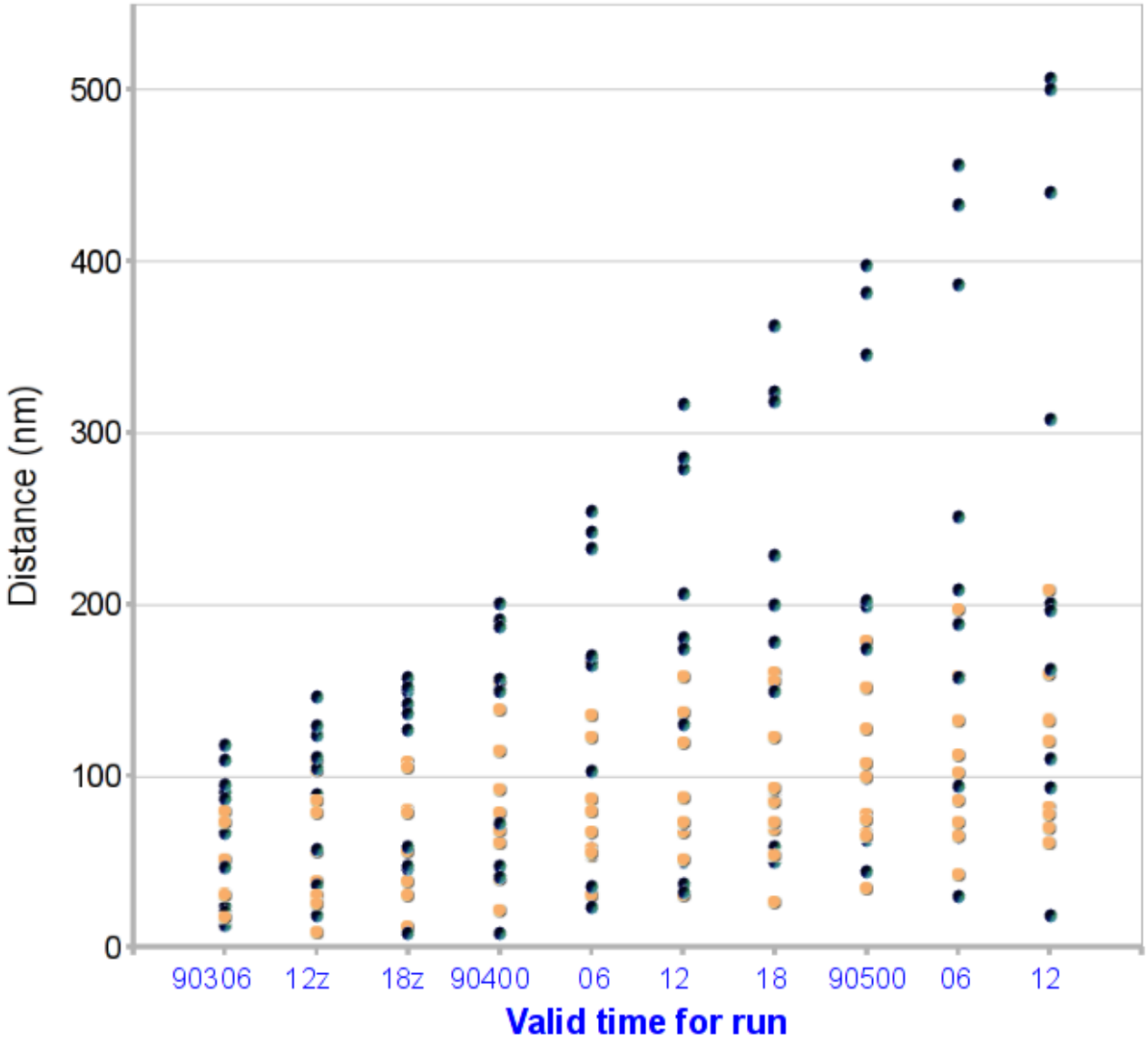


Figure 6.1.6: Distance (nm) between AOM forecasts of Felix storm center initialized at the times listed on Fig. 5.1 and valid at the September 2007 times listed on the x-axis. Black (yellow) is the low- (high-) resolution forecast.

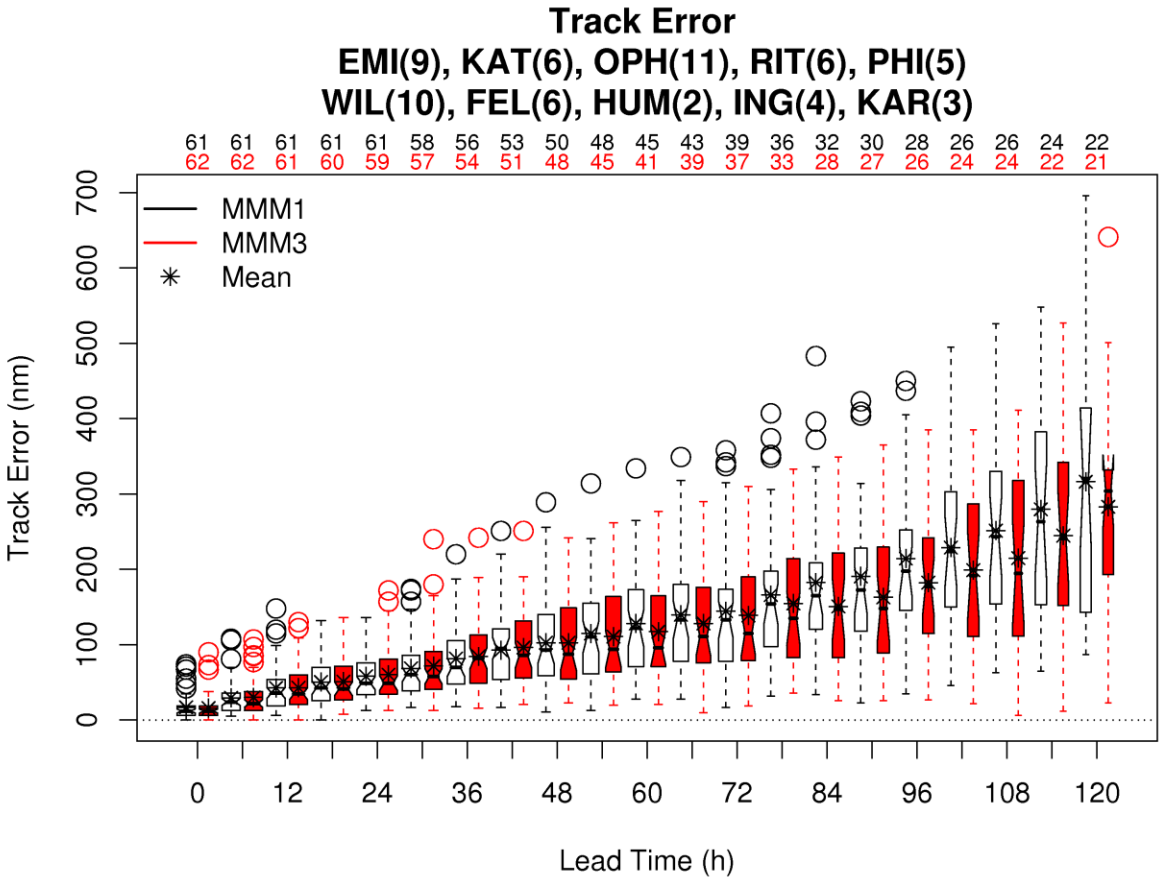
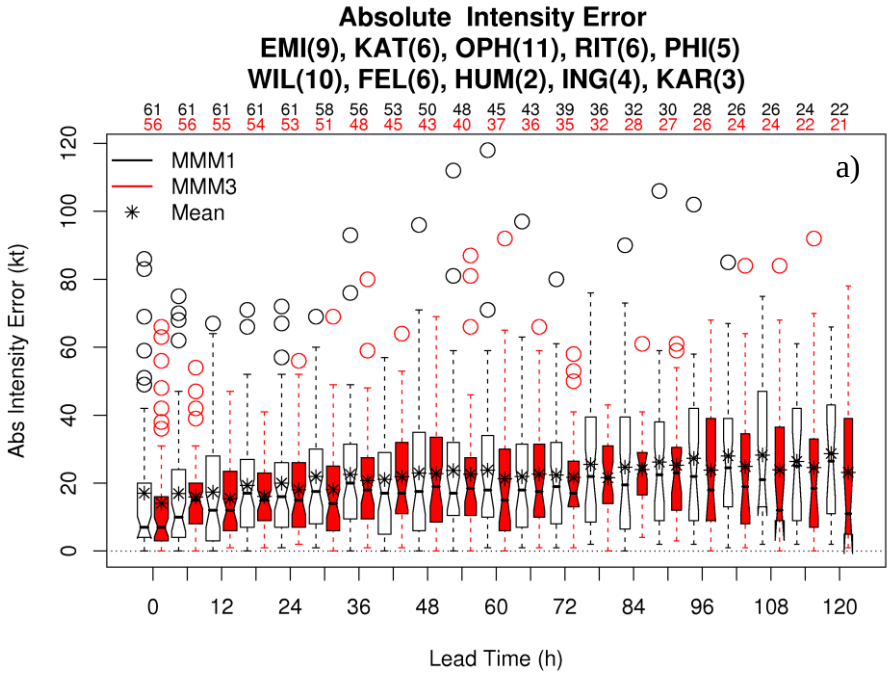


Figure 6.2.1: Same as Fig. 6.1.1 except for MMM model configurations.



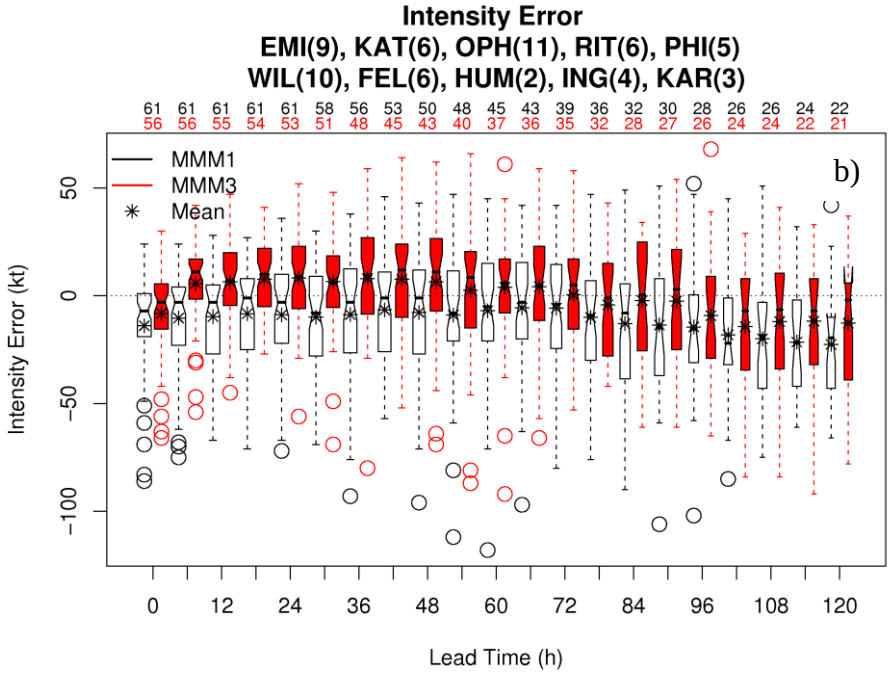


Figure 6.2.2: Same as Fig. 6.1.2, except MMM configurations.

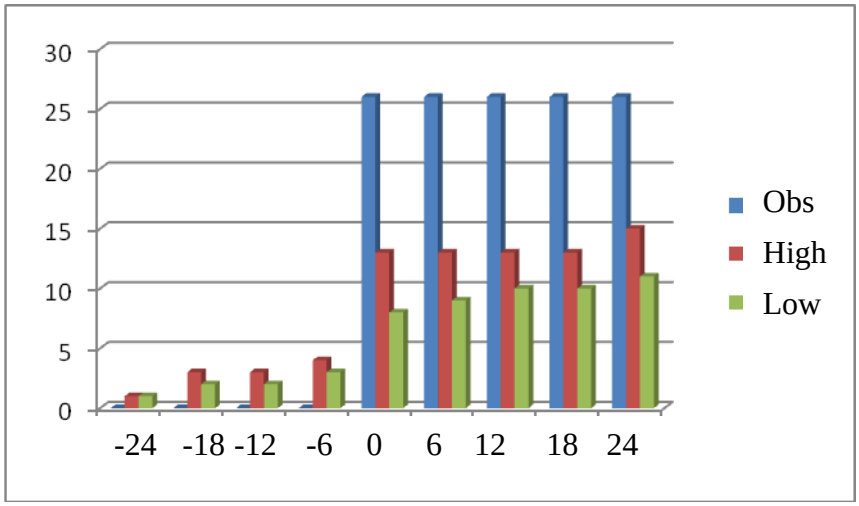


Figure 6.2.3: Same as Fig. 6.1.3, except for MMM configurations.

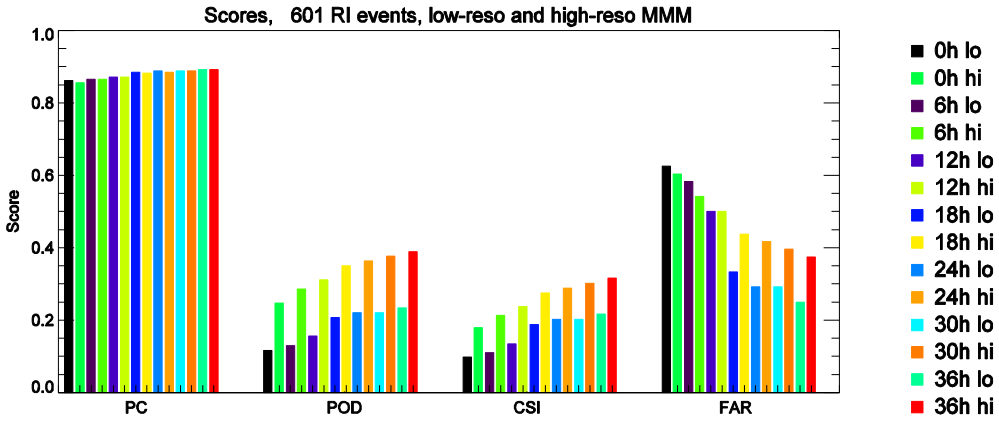
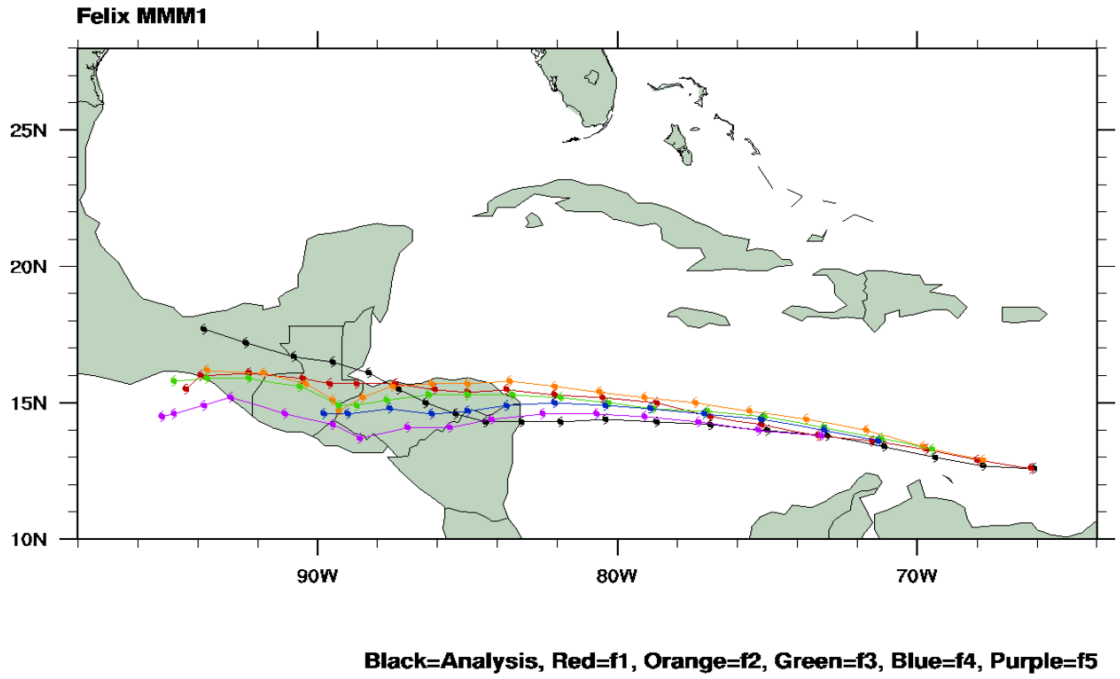


Figure 6.2.4: Same as Fig. 6.1.4, except for MMM configurations.



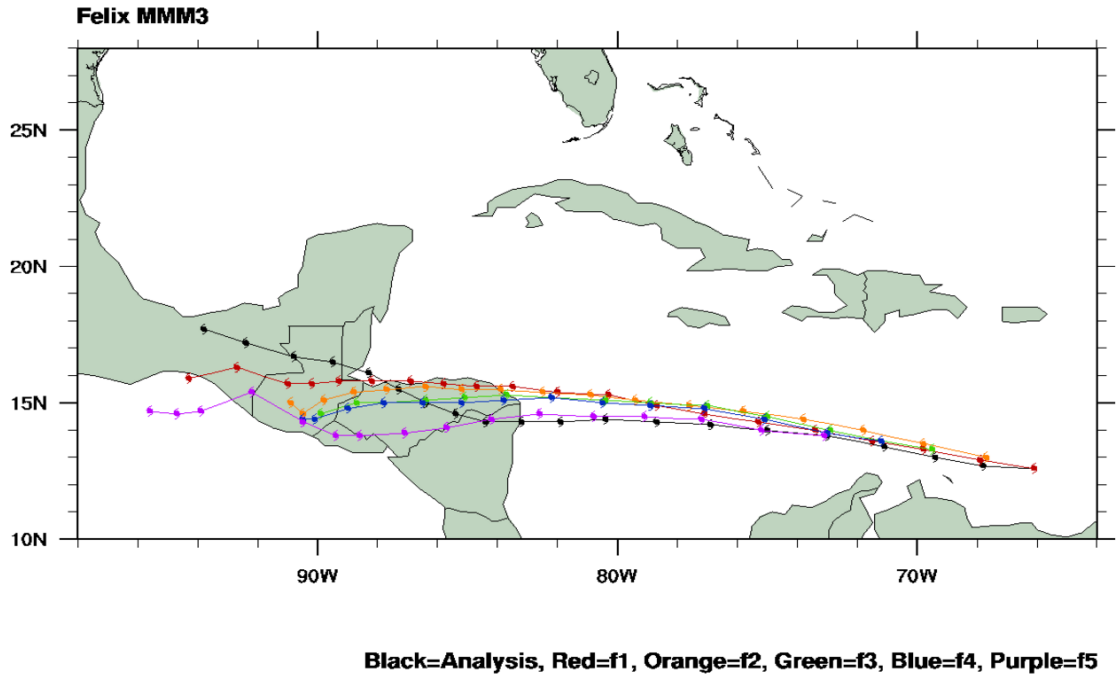


Figure 6.2.5: Same as Fig. 6.1.5, except for MMM.

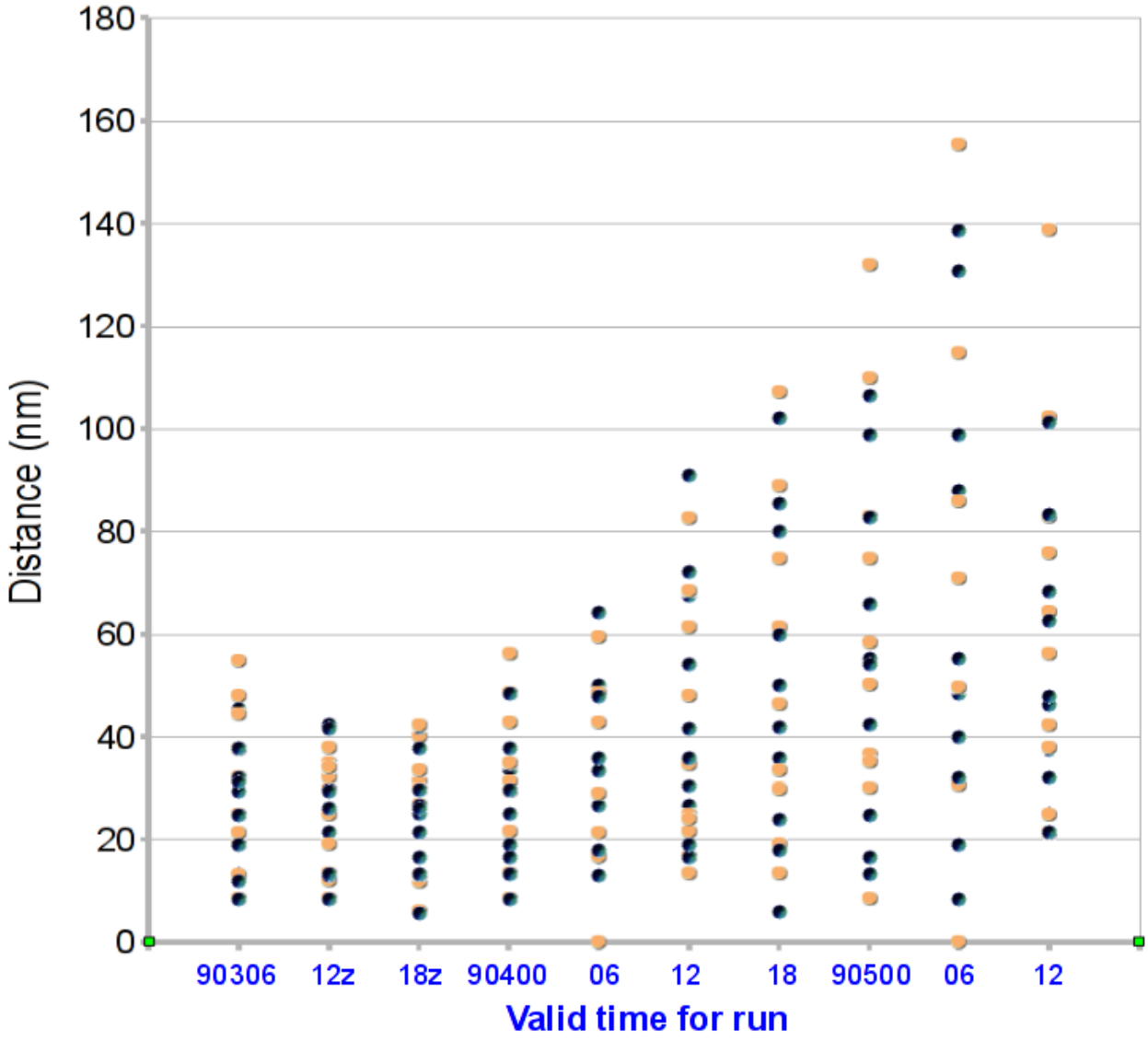


Figure 6.2.6: Same as Fig. 6.1.6, except for MMM.

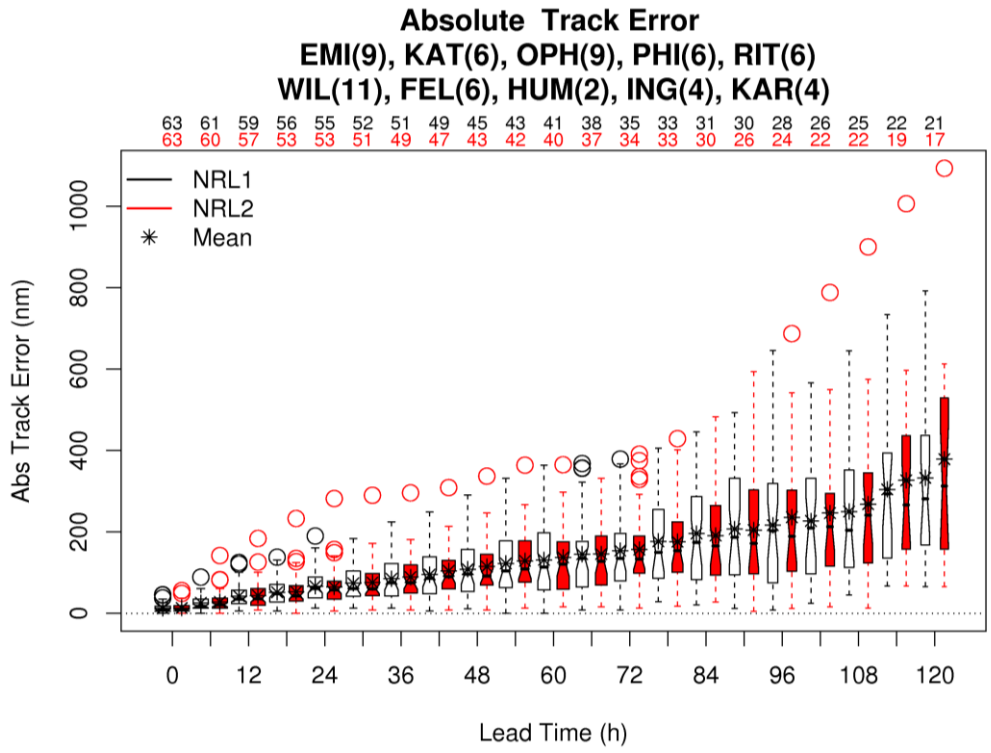


Figure 6.3.1: Same as Fig. 6.1.1 except for NRL configurations.

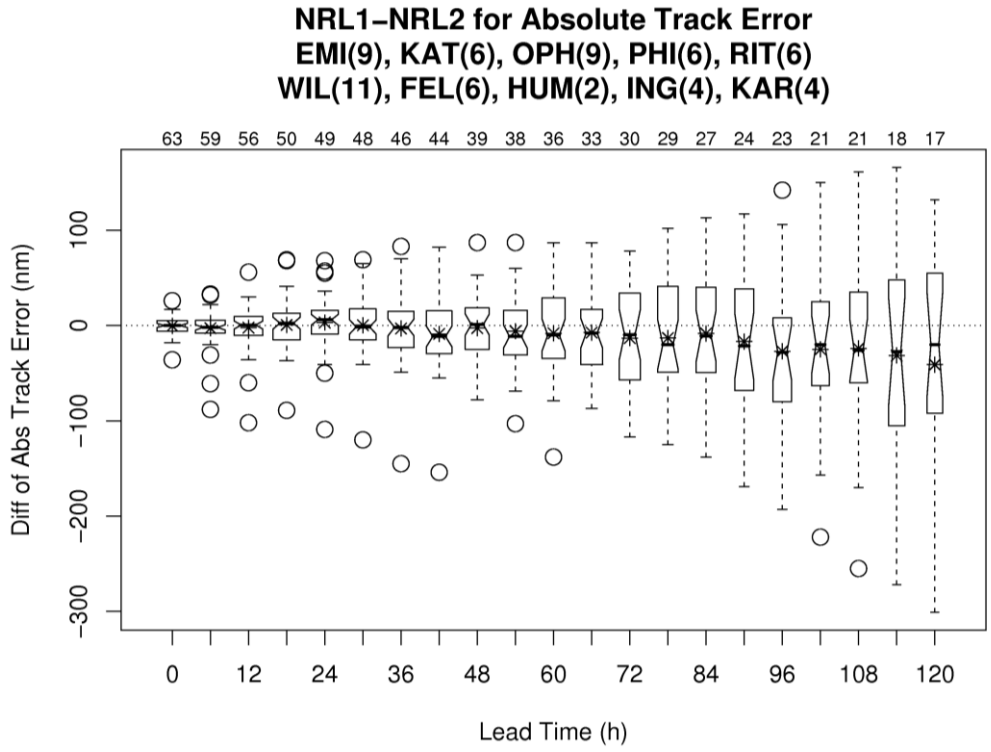


Figure 6.3.2: Distributions of the pairwise differences between the track errors for the high- and low-resolution configuration of NRL with respect to lead time.

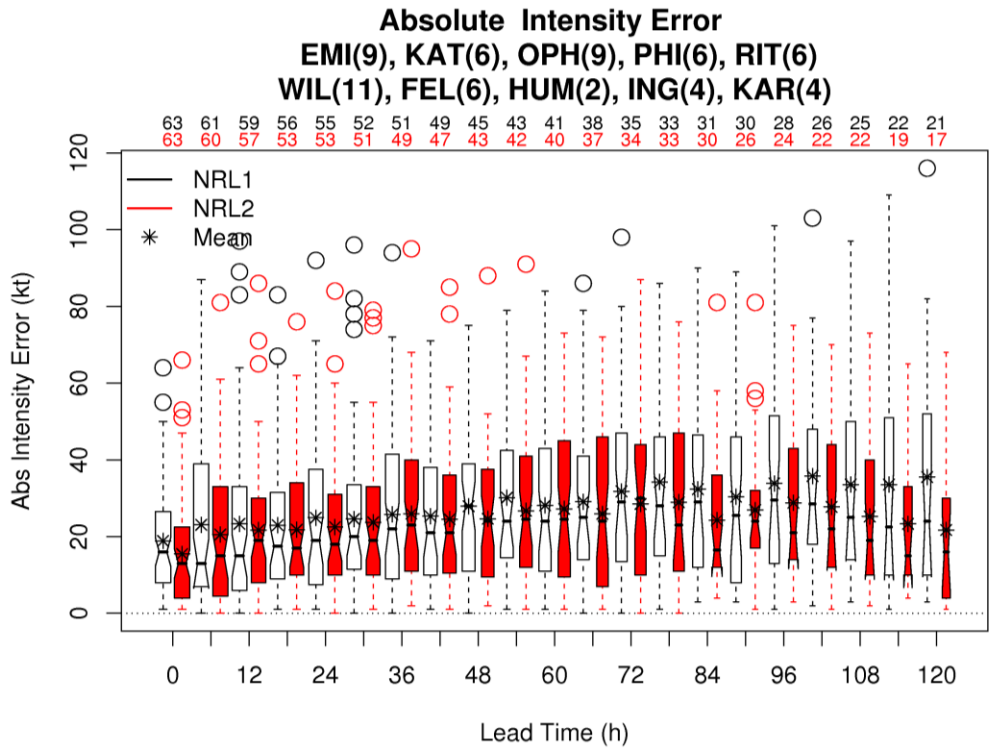


Figure 6.3.3: Absolute intensity error distributions with respect to lead time for the high- and low-resolution NRL model configurations.

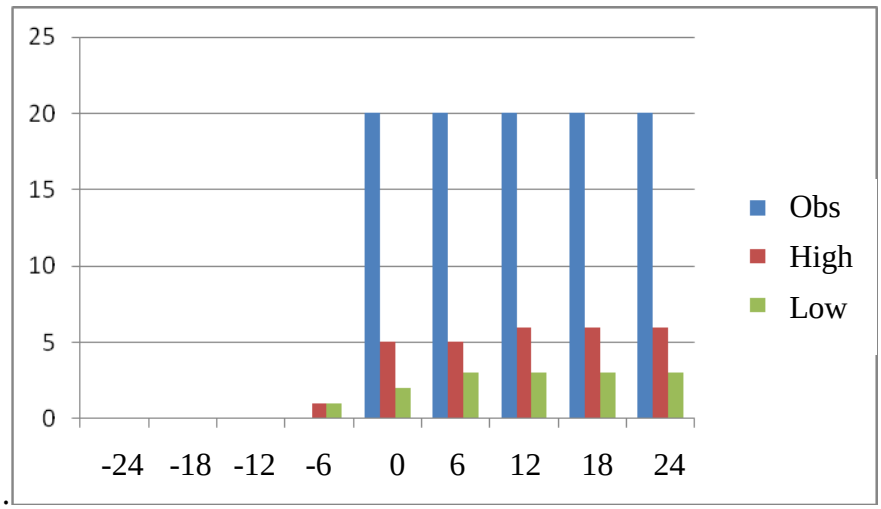


Figure 6.3.4: Same as Fig. 6.1.3 except for NRL configurations.

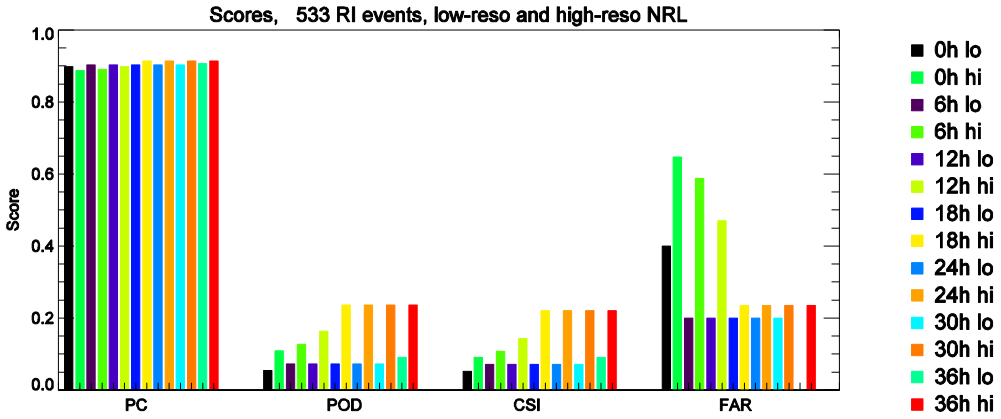
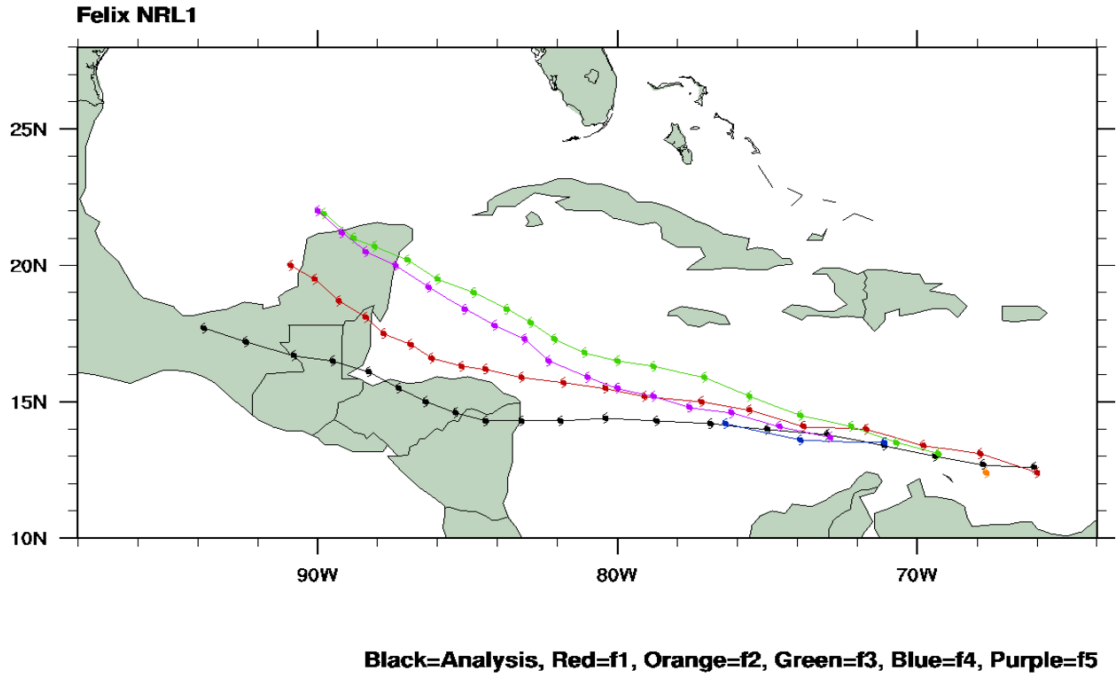


Figure 6.3.5: Same as Fig. 6.1.4 except for NRL configurations.



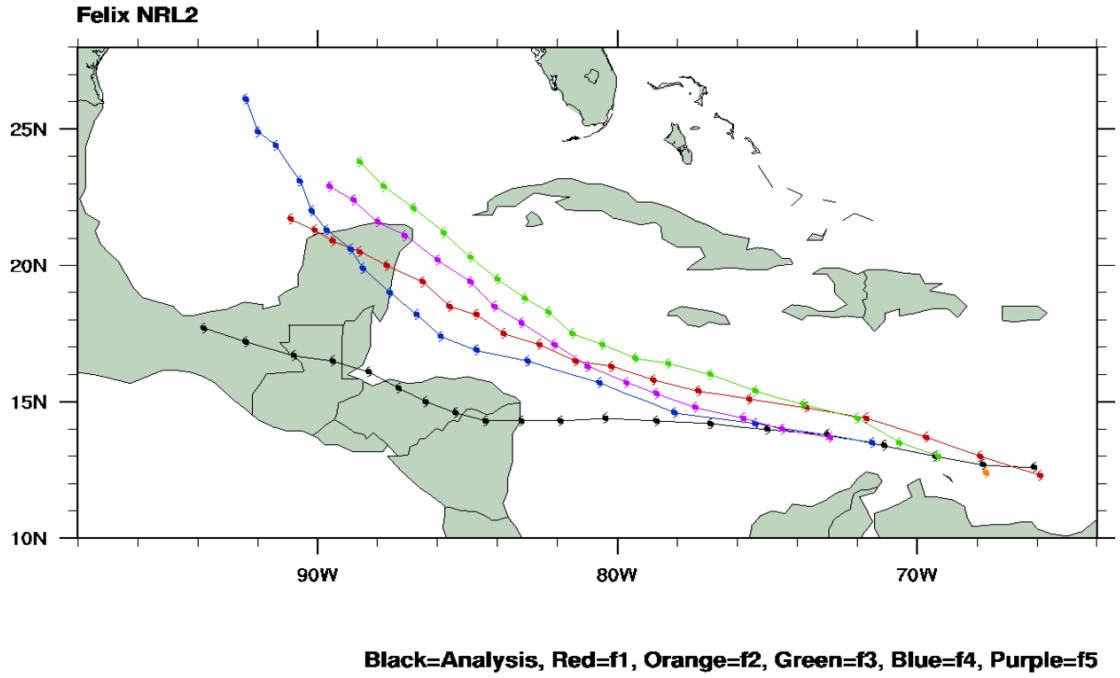


Figure 6.3.6: Same as Fig. 6.1.5, except for NRL.

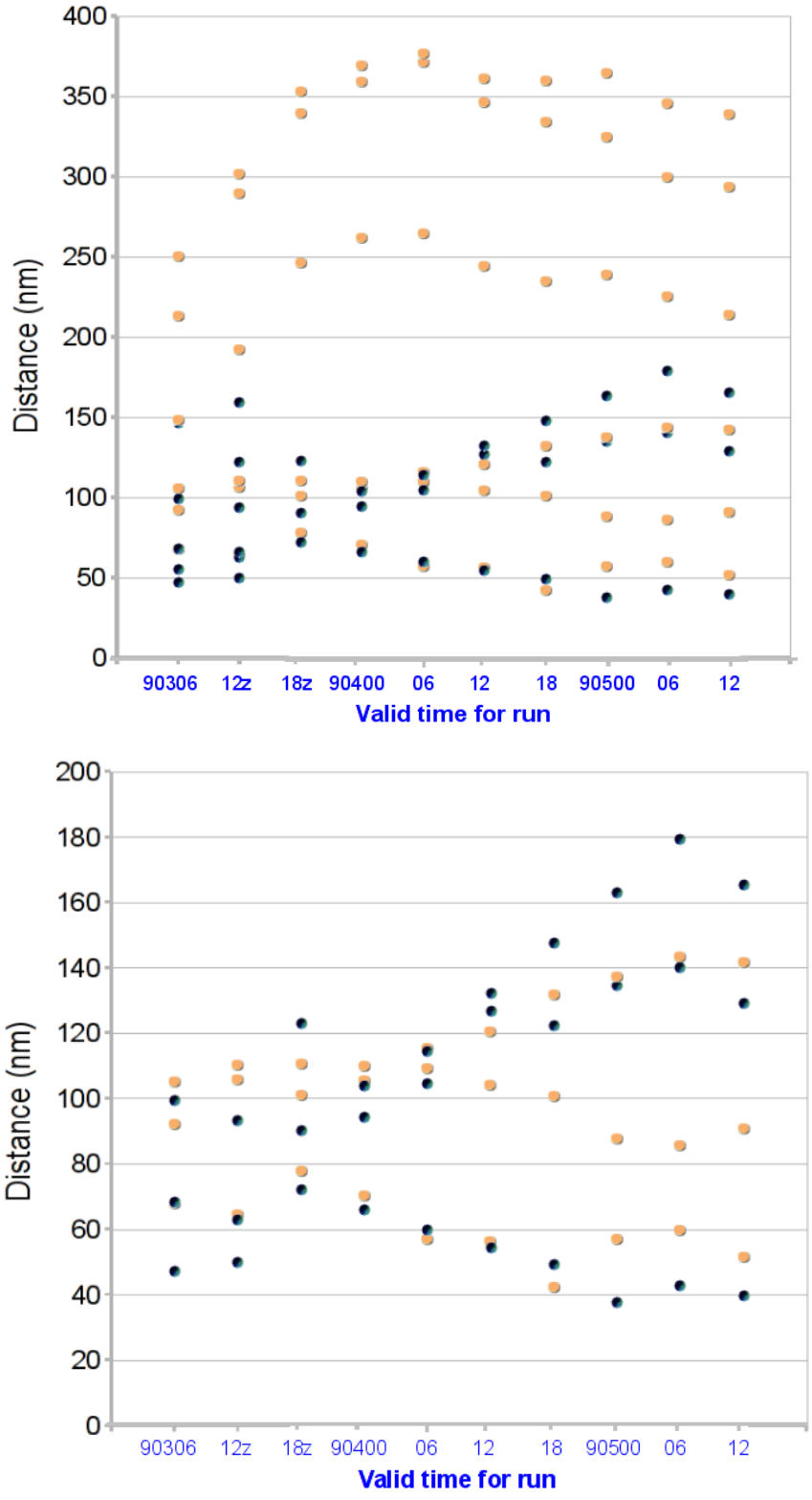


Figure 6.3.7: Same as Fig. 6.1.6, except for NRL and a) contains all forecasts and b) only those for which both low- and high-resolution runs are complete.

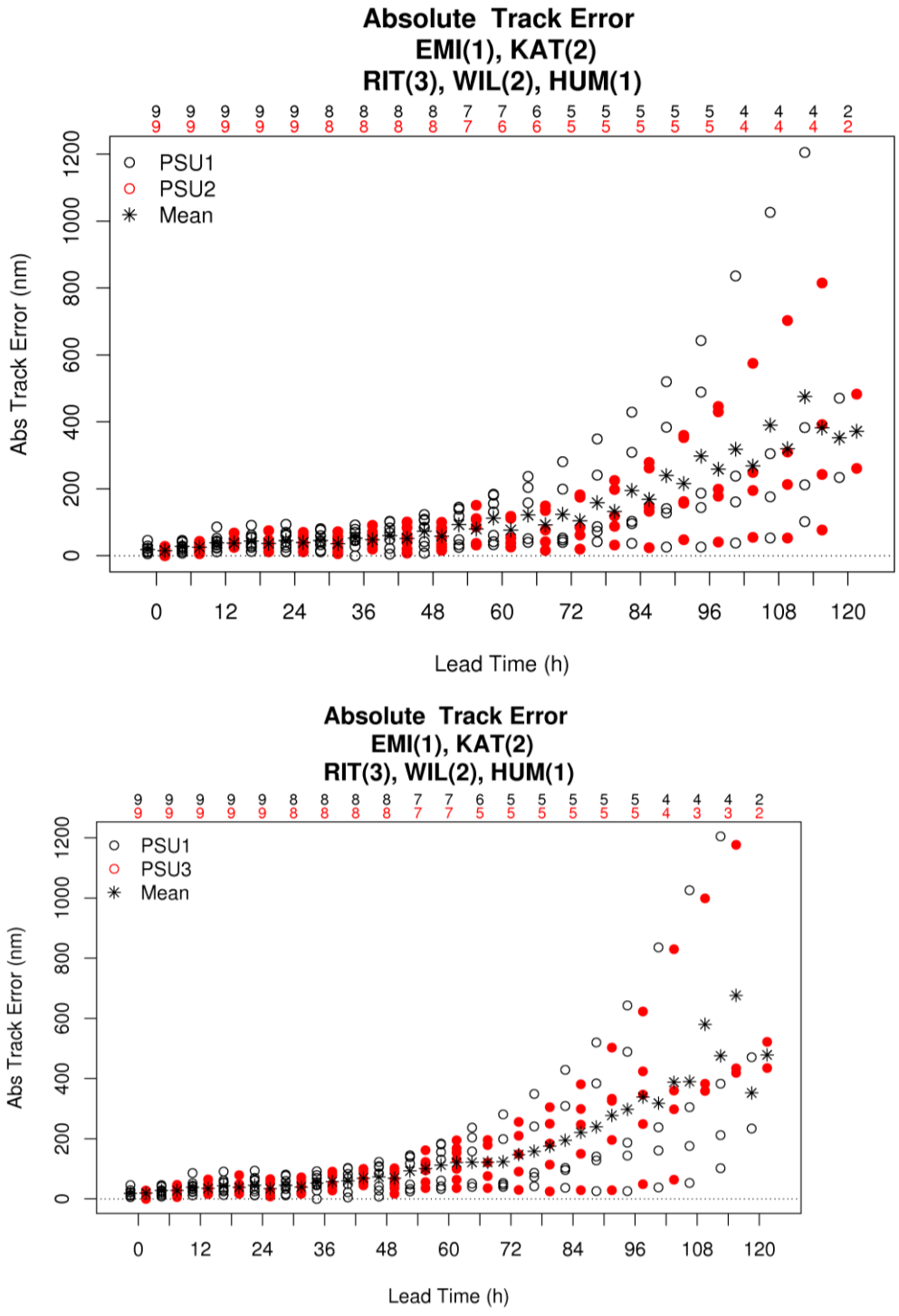


Figure 6.4.1: Scatter plots of PSU track error with respect to lead time for a) low- and intermediate-resolutions configurations, and b) low- and high-resolution configurations.

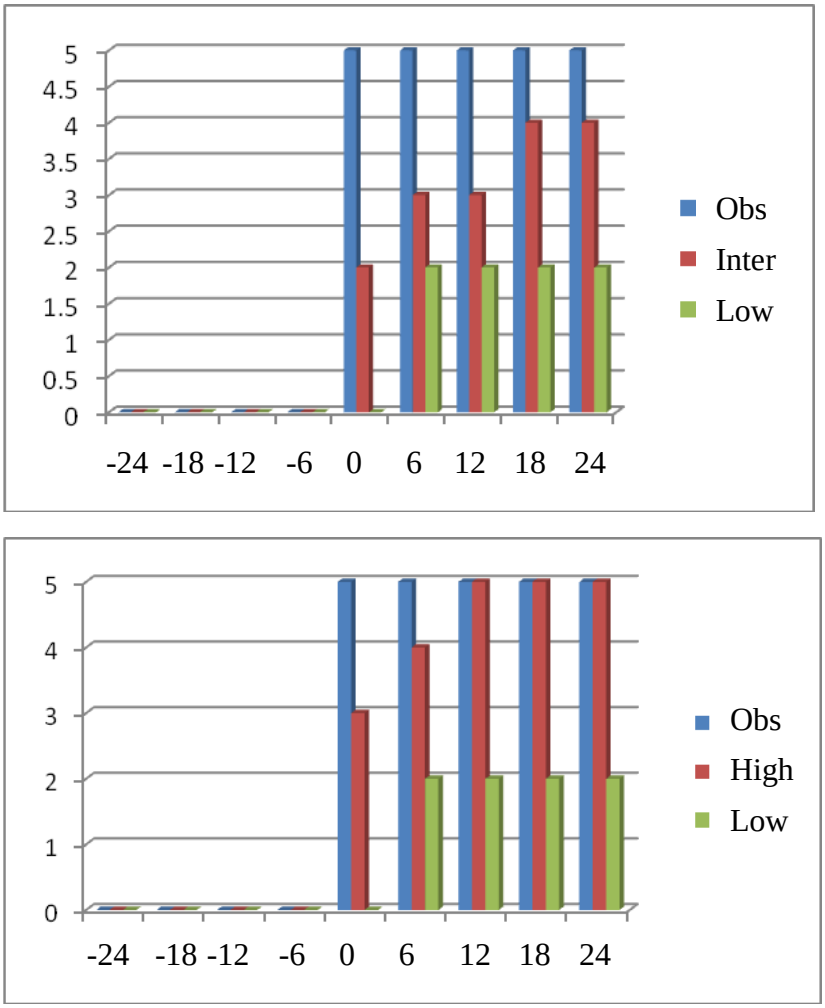


Figure 6.4.2: Same as Fig. 6.1.3 except for PSU configurations (top panel – low- and intermediate-resolutions, bottom panel – low- and high-resolutions)

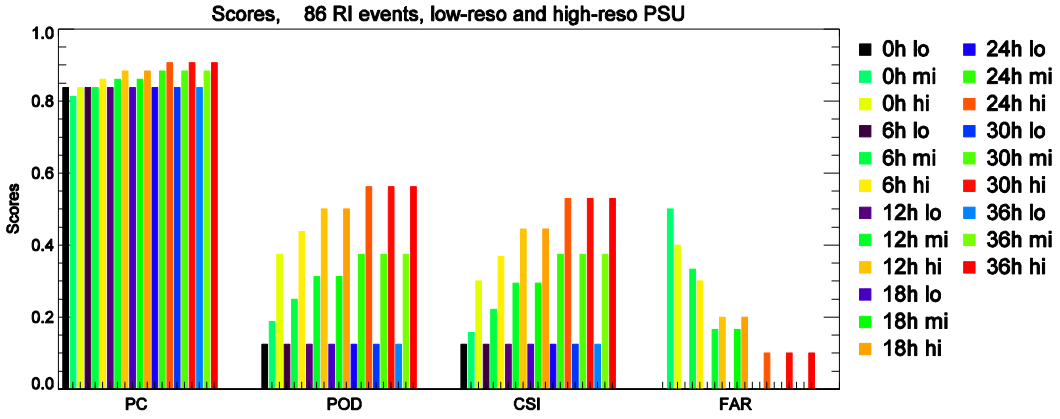


Figure 6.4.3: Same as Fig. 6.1.4 except for PSU configurations

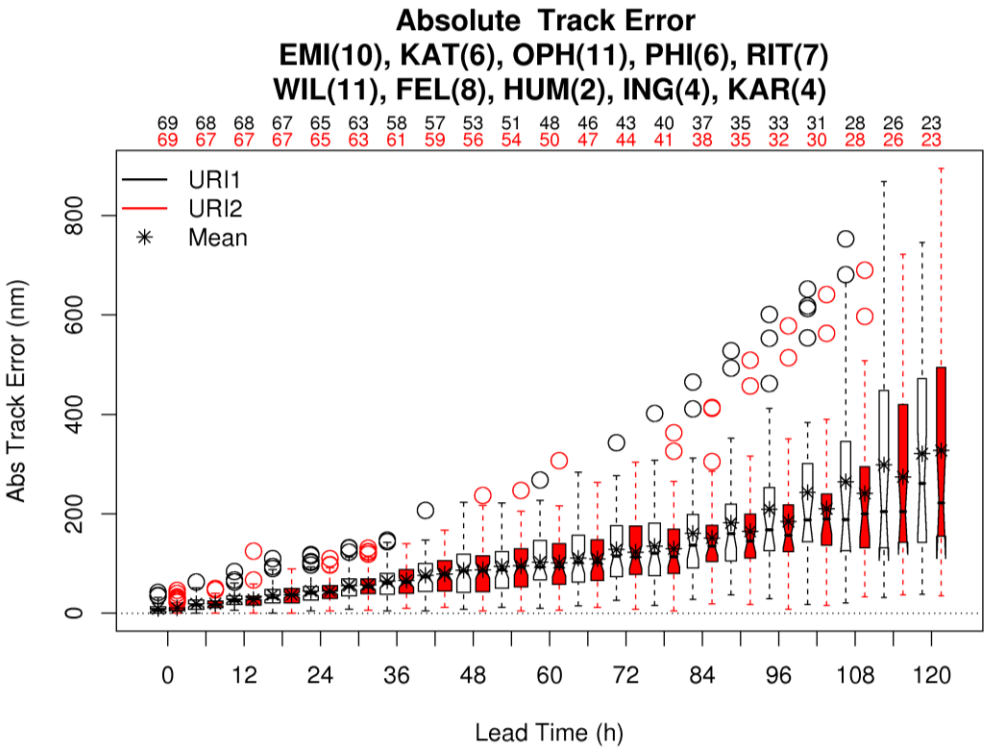


Figure 6.5.1: Same as Fig. 6.1.1 except for URI configurations.

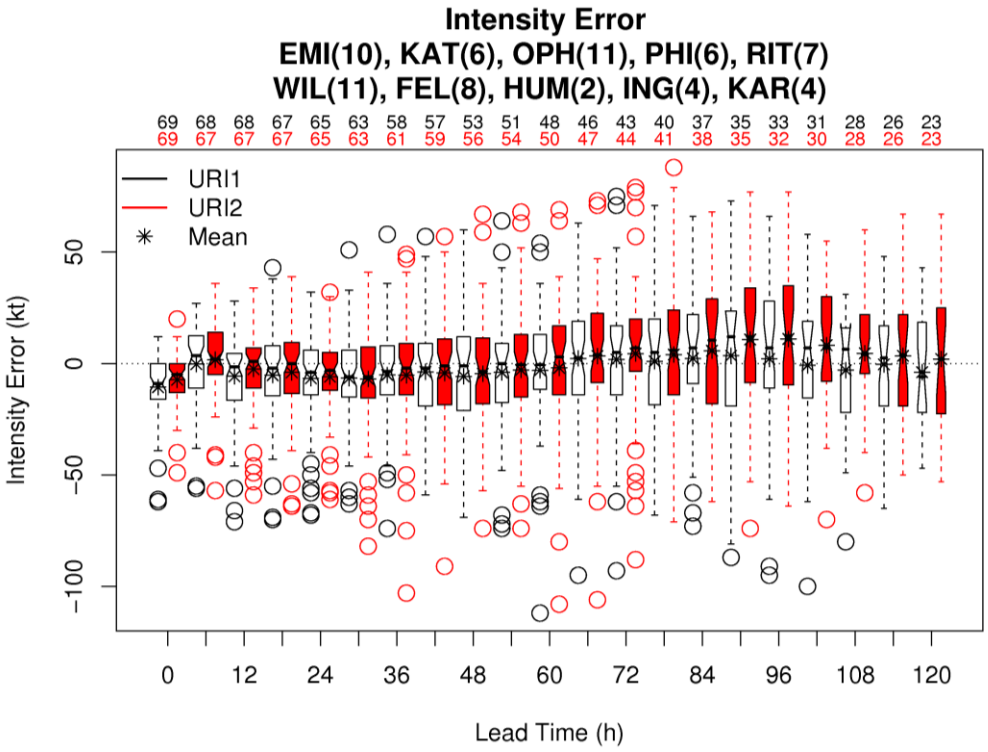


Figure 6.5.2: Intensity error distributions with respect to lead time for the high- and low-resolution URI model configurations.

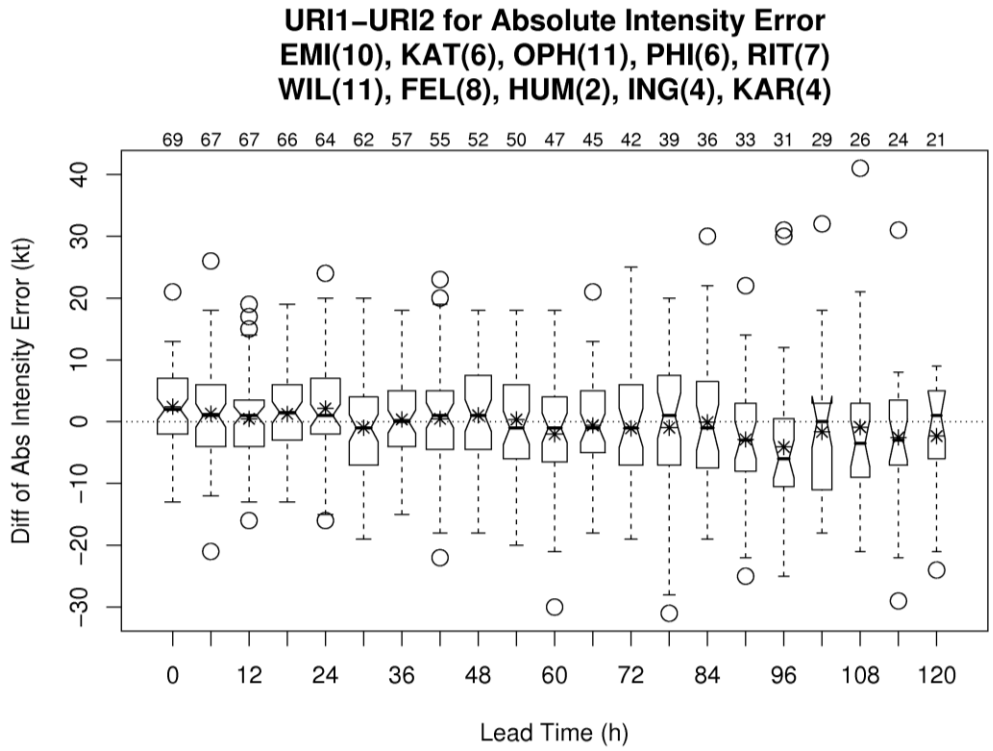


Figure 6.5.3: Distributions of the pairwise differences between the absolute intensity errors for the high- and low-resolution configuration of URI with respect to lead time.

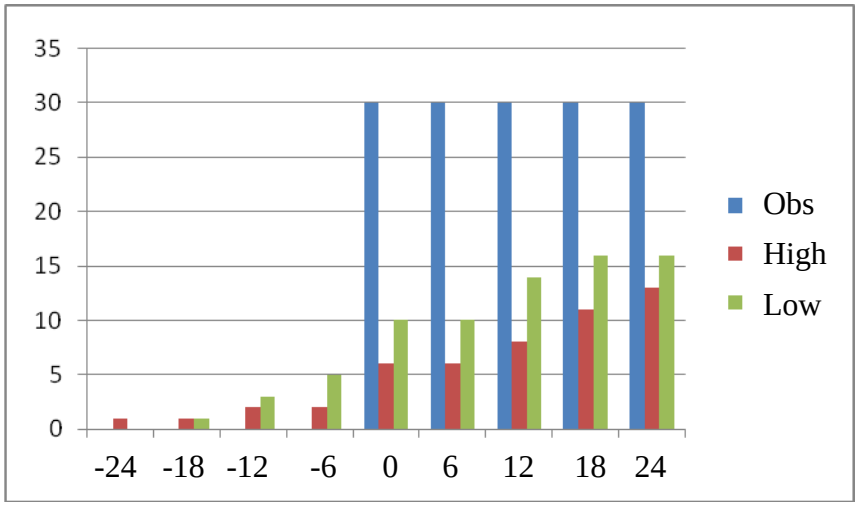


Figure 6.5.4: Same as Fig. 6.1.3 except for URI configurations.

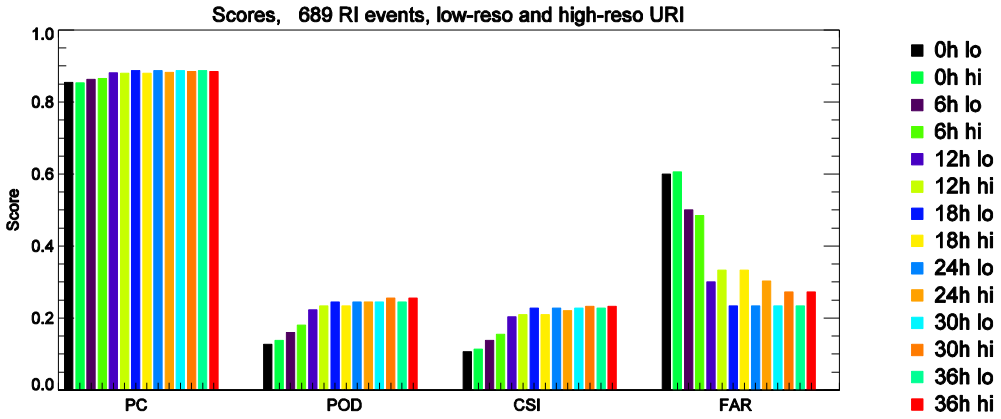
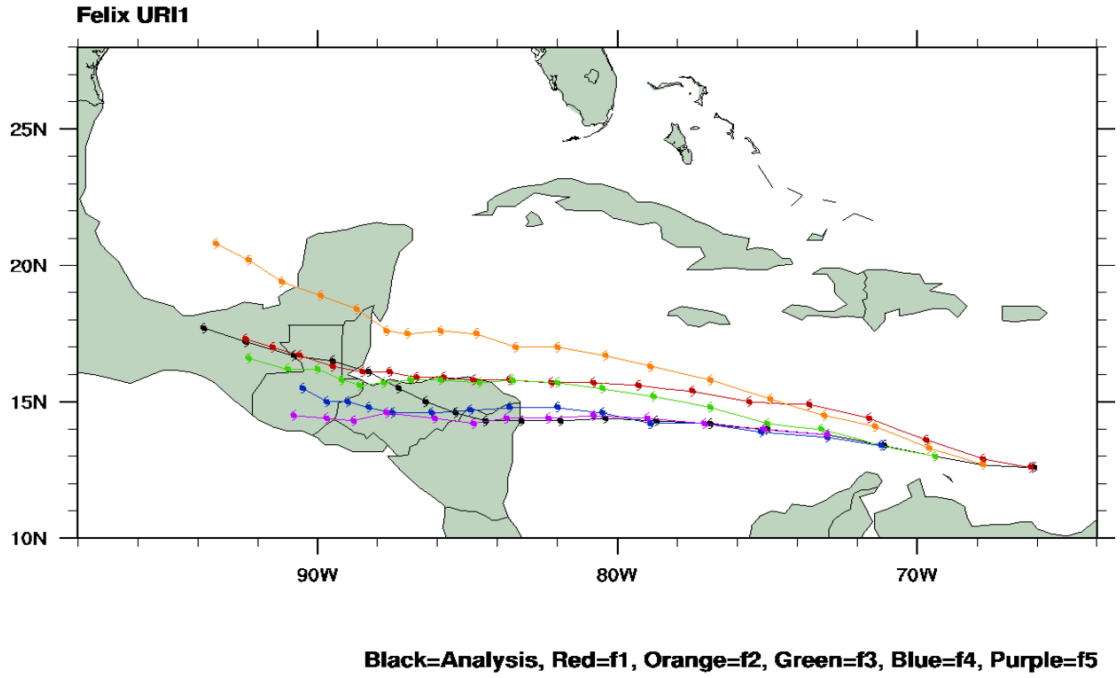


Figure 6.5.5: Same as Fig. 6.1.3 except for URI configurations.



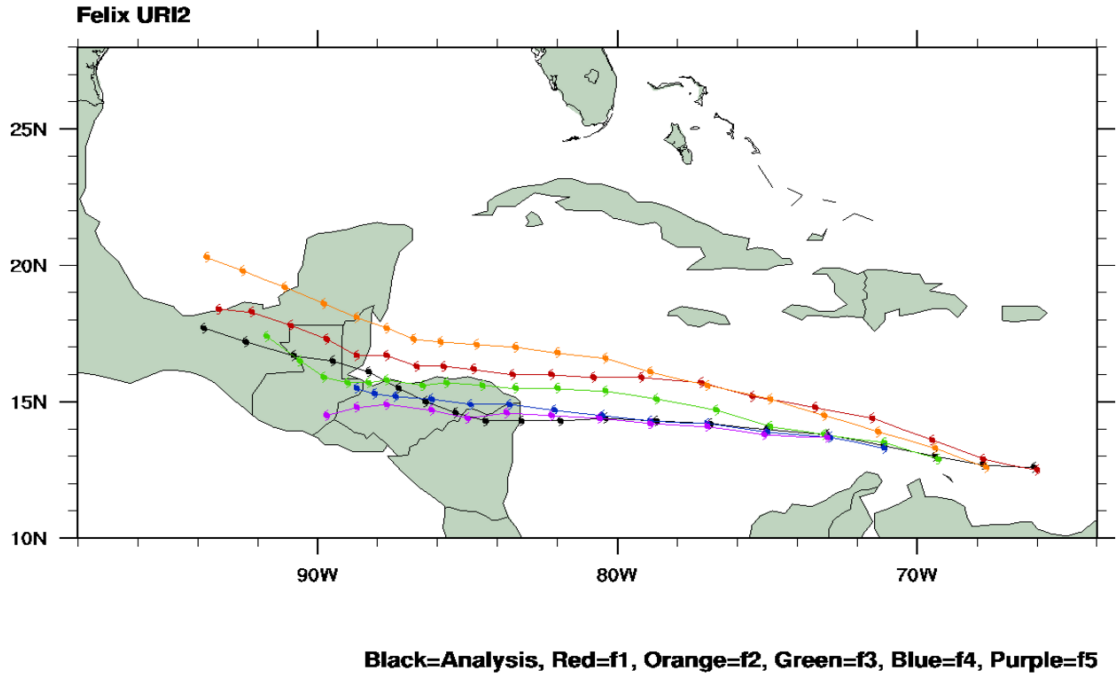


Figure 6.5.6: Same as Fig. 6.1.5, except for URI.

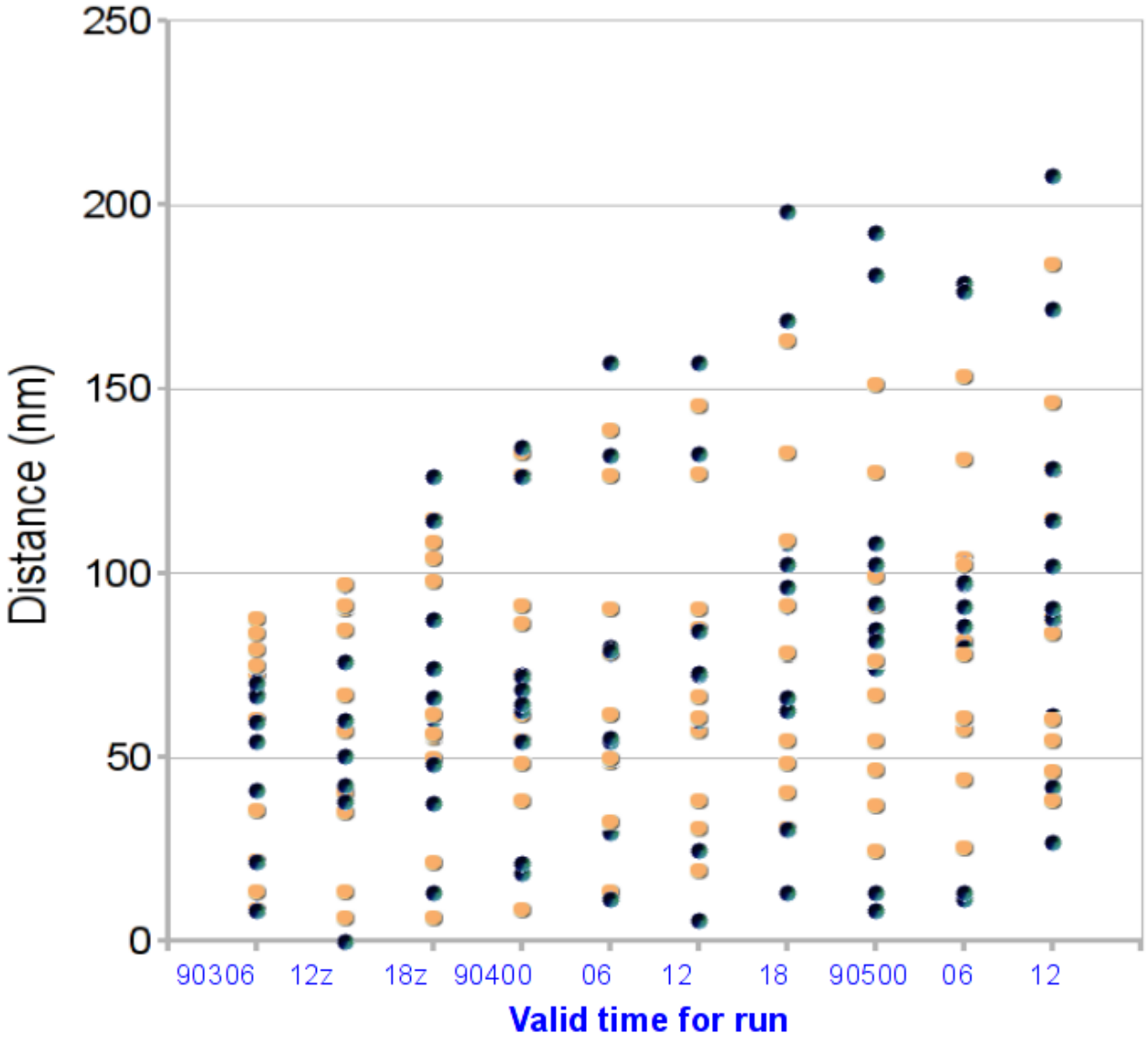


Figure 6.5.7: Same as Fig. 6.1.6, except for URI.

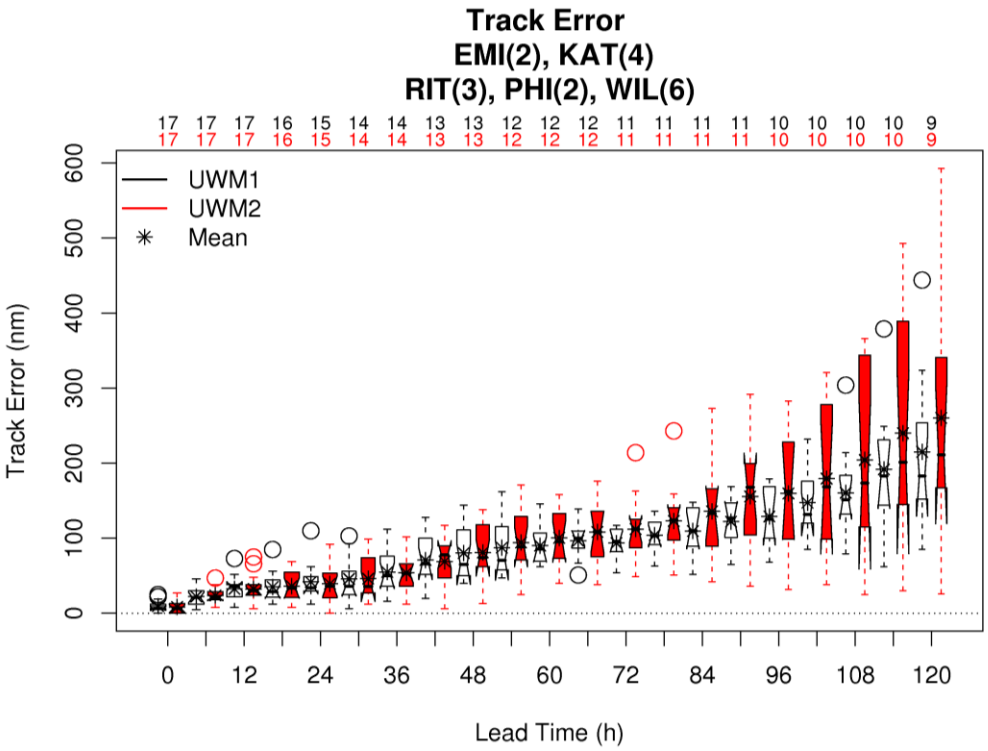


Figure 6.6.1: Same as Fig. 6.1.1, except for UWM low- and intermediate-resolutions.

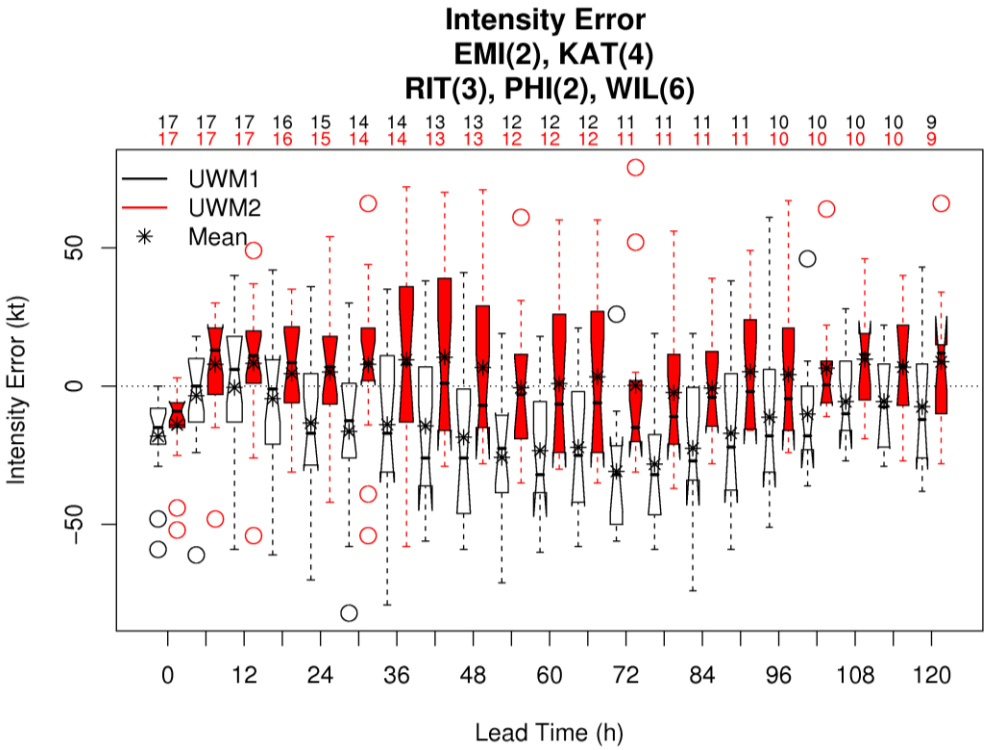


Figure 6.6.2: Same as Fig. 6.5.2, except for UWM low- and intermediate-configurations.

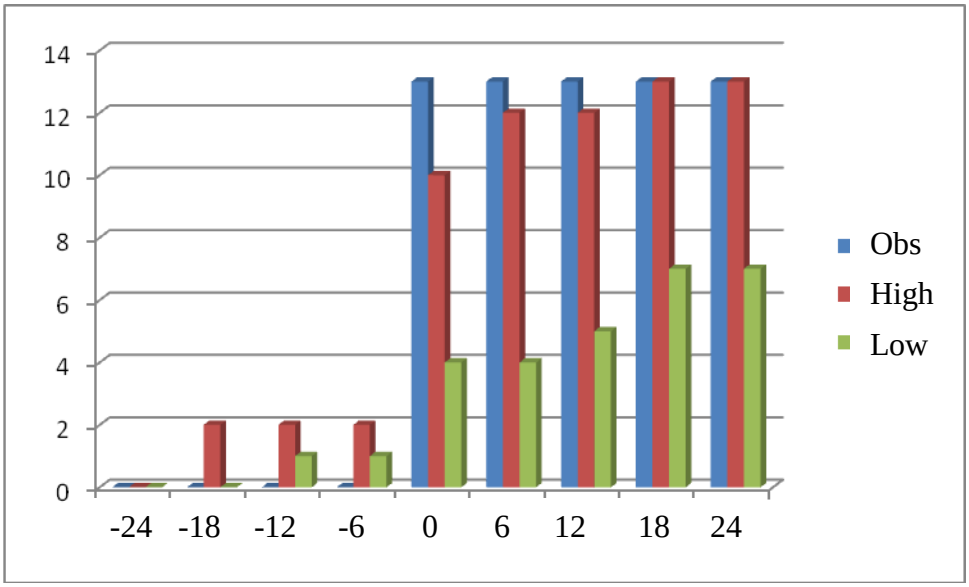


Figure 6.6.3: Same as Fig. 6.1.3, except for UWM low- and intermediate-resolutions.

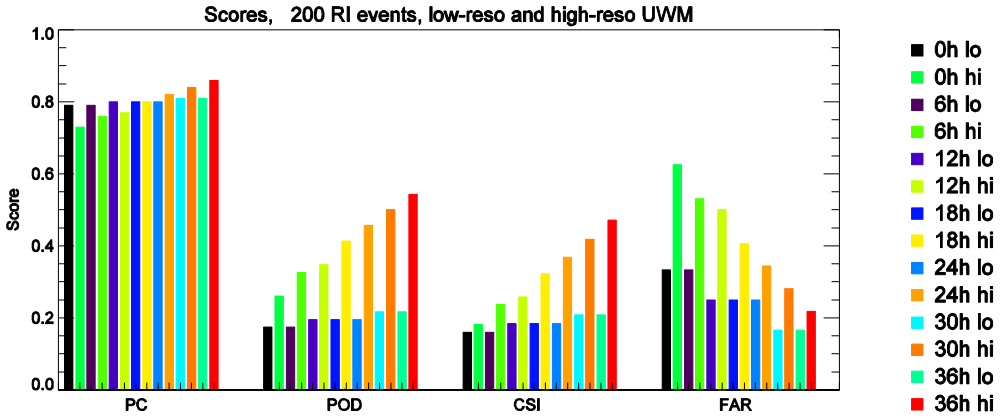


Figure 6.6.4: Same as Fig. 6.1.4, except for UWM low- and intermediate-resolutions.

Appendix A: List of workshop participants and their affiliation

Participant	Affiliation	March 2008	May 2009
Fred Toepfer	NOAA/HFIP	X	X
Frank Marks	NOAA/AOML	X	X
S. Gopalakrishnan	NOAA/AOML	X	X
Robert Rogers	NOAA/AOML	X	X
Xuejin Zhang	NOAA/AOML	X	X
Kevin Yeh	NOAA/AOML	X	
Kathryn Sellwood	NOAA/AOML		X
Eric Uhlhorn	NOAA/AOML		X
Bill Read	NOAA/NHC	X	
Mike Fiorino	NOAA/NHC	X	X
James Franklin	NOAA/NHC	X	X
Ahsha Tribble	NOAA/NHC	X	
Jack Beven	NOAA/NHC	X	X
Richard Pasch	NOAA/NHC		X
Edward Rappaport	NOAA/NHC		X
Naomi Surgi	NOAA/NCEP/EMC	X	X
Steve Lord	NOAA/NCEP/EMC	X	
Hyun-Sook Kim	NOAA/NCEP/EMC		X
Young Kwon	NOAA/NCEP/EMC		X
Vijay Tallapragada	NOAA/NCEP/EMC		X
Robert Tuleya	NOAA/SAIC		X
Mark DeMaria	NOAA/CIRA	X	X
Kate Maclay	CSU/CIRA		X
Jian-Wen Bao	NOAA/ESRL	X	X
Ligia Bernardet	NOAA/ESRL/DTC	X	X
Shaowu Bao	NOAA/ESRL/DTC		X
Morris Bender	NOAA/GFDL	X	
Tim Marchok	NOAA/GFDL	X	X
Isaac Ginis	University of Rhode Island	X	X
Richard Yablonsky	University of Rhode Island		X
Bob Gall	NCAR/RAL/DTC – NOAA/HFIP	X	X
Rich Wagoner	NCAR/RAL	X	
Louisa Nance	NCAR/RAL/DTC	X	X
Barb Brown	NCAR/RAL/DTC	X	X
Bill Kuo	NCAR/RAL/DTC		X
Greg Holland	NCAR/ESSL/MMM	X	
Chris Davis	NCAR/ESSL/MMM	X	X
Bo-Wen Shen	NASA/GSFC	X	
Melinda Peng	DOD/NRL	X	
Hao Jin	DOD/NRL		X
Shuyi Chen	University of Miami/RSMAS	X	
Nick Shay	University of Miami/RSMAS		X
Greg Tripoli	University of Wisconsin-Madison	X	X
Mike Montgomery	Navy Postgraduate School	X	
Fuquing Zhang	Pennsylvania State University	X	X
Roger Smith	University of Munich	X	
T. N. Krishnamurti	Florida State University		X
Daniel Melendez	NOAA/NWS	X	

Appendix B: Selected test cases and data inventories

Table B1: Data inventory for NRL, URI, MMM and AOML. A check mark indicates gridded data was received containing all requested lead times. An empty cell indicates no gridded data was received for that particular case. A numerical entry indicates the longest lead time for which gridded data was received for those cases where the delivered lead times fell short of that requested.

Storm	# of Cases	Forecast Date	Forecast Time	Hours w/ track	NRL1	NRL2	NRL5	URI1	URI2	MMM1	MMM3	MMM4	AOM1	AOM2	AOM5	AOM6
Wilma	**11**				**11**	**11**	**11**	**11**	**11**	**11**	**11**	**11**	**11**	**11**	**11**	**11**
		10/16/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/17/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/18/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/19/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/19/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/20/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/21/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/22/2005	0	114	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/23/2005	0	90	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/24/2005	0	66	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		10/25/2005	0	42	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Philippe	**6**				**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**
		9/17/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/18/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/19/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/20/2005	12	90	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/21/2005	12	66	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/22/2005	12	42	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Felix	**8**				**6**	**6**	**6**	**8**	**8**	**8**	**8**	**8**	**8**	**8**	**8**	**8**
		8/31/2007	12	126				✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/1/2007	12	126				✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/2/2007	0	114	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/2/2007	6	108	✓	✓	✓	✓	✓	✓	✓	75.5	✓	✓	✓	✓
		9/2/2007	12	102	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/2/2007	18	96	✓	✓	✓	89	✓	✓	✓	✓	✓	✓	✓	✓
		9/3/2007	0	90	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Storm	# of Cases	Forecast Date	Forecast Time	Hours w/ track	NRL1	NRL2	NRL5	URI1	URI2	MMM1	MMM3	MMM4	AOM1	AOM2	AOM5	AOM6
		9/3/2007	12	78	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rita	**7**				**6**	**6**	**6**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**
		9/18/2005	0	126				✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/19/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/20/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/21/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/22/2005	0	102	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/23/2005	0	78	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/24/2005	0	54	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Karen	**4**				**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**
		9/25/2007	0	108	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/26/2007	0	84	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/27/2007	0	60	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/28/2007	0	36	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Katrina	**6**				**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**
		8/24/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		8/25/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		8/26/2005	0	126	✓	✓	✓	✓	✓	125	✓	✓	✓	✓	✓	✓
		8/27/2005	0	102	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		8/28/2005	0	78	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		8/29/2005	0	54	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Humberto	**2**				**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
		9/12/2007	12	48	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/13/2007	0	36	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ingrid	**4**				**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**
		9/12/2007	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/13/2007	12	120	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/14/2007	12	96	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/15/2007	12	72	✓	✓	✓	49	59	✓	✓	✓	✓	✓	✓	✓
Emily	**10**				**9**	**9**	**9**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**
		7/11/2005	0	126				✓	69	✓	✓	✓	✓	✓	✓	✓

Storm	# of Cases	Forecast Date	Forecast Time	Hours w/ track	NRL1	NRL2	NRL5	URI1	URI2	MMM1	MMM3	MMM4	AOM1	AOM2	AOM5	AOM6
		7/12/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		7/13/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		7/14/2005	0	126	✓	✓	✓	✓	88.5	✓	✓	✓	✓	✓	✓	✓
		7/15/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		7/16/2005	0	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		7/17/2005	0	108	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		7/18/2005	0	84	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		7/19/2005	0	60	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		7/20/2005	0	36	✓	✓	✓	0	0	✓	✓	✓	✓	✓	✓	✓
Ophelia	**11**				**9**	**9**	**9**	**11**	**11**	**11**	**11**	**11**	**11**	**11**	**11**	**11**
		9/6/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/7/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/8/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/9/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/10/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/11/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/12/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/13/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/14/2005	12	126	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		9/15/2005	12	126				106	108	✓	✓	✓	✓	✓	✓	✓
		9/16/2005	12	126				86	86	✓	✓	✓	✓	✓	✓	✓
Total	**69**				**63**	**63**	**63**	**69**	**69**	**69**	**69**	**69**	**69**	**69**	**69**	**69**

Table B2: Same as Table B3 except for UWM and PSU.

Storm	# of Cases	Forecast Date	Forecast Time	Hours w/ track	UWM1	UWM2	UWM3	UWM4	UWM5	PSU1	PSU2	PSU3	PSU4	PSU5
Wilma	**11**				**6**	**6**	**0**	**0**	**6**	**2**	**2**	**2**	**2**	**2**
		10/16/2005	0	126	✓	✓			✓					
		10/17/2005	0	126	✓	✓			✓					
		10/18/2005	0	126	✓	✓			✓					
		10/19/2005	0	126	✓	✓			✓					

Storm	# of Cases	Forecast Date	Forecast Time	Hours w/ track	UWM1	UWM2	UWM3	UWM4	UWM5	PSU1	PSU2	PSU3	PSU4	PSU5
		10/19/2005	12	126	✓	✓			✓					
		10/20/2005	0	126										
		10/21/2005	0	126						✓	✓	✓	✓	✓
		10/22/2005	0	114										
		10/23/2005	0	90						✓	✓	✓	✓	✓
		10/24/2005	0	66										
		10/25/2005	0	42	✓	✓			✓					
Philippe	**6**				**2**	**2**	**0**	**0**	**2**	**0**	**0**	**0**	**0**	**0**
		9/17/2005	12	126										
		9/18/2005	12	126										
		9/19/2005	12	126										
		9/20/2005	12	90										
		9/21/2005	12	66	✓	✓			✓					
		9/22/2005	12	42	✓	✓			✓					
Felix	**8**				**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
		8/31/2007	12	126										
		9/1/2007	12	126										
		9/2/2007	0	114										
		9/2/2007	6	108										
		9/2/2007	12	102										
		9/2/2007	18	96										
		9/3/2007	0	90										
		9/3/2007	12	78										
Rita	**7**				**3**	**3**	**0**	**0**	**3**	**3**	**3**	**3**	**3**	**3**
		9/18/2005	0	126										
		9/19/2005	0	126	✓	✓			✓					
		9/20/2005	0	126	✓	✓			✓	✓	✓	✓	✓	✓
		9/21/2005	0	126	✓	✓			✓	✓	✓	✓	✓	✓
		9/22/2005	0	102						✓	✓	✓	✓	✓
		9/23/2005	0	78										
		9/24/2005	0	54										
Karen	**4**				**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Storm	# of Cases	Forecast Date	Forecast Time	Hours w/ track	UWM1	UWM2	UWM3	UWM4	UWM5	PSU1	PSU2	PSU3	PSU4	PSU5
		9/25/2007	0	108										
		9/26/2007	0	84										
		9/27/2007	0	60										
		9/28/2007	0	36										
Katrina	**6**				**4**	**4**	**2**	**2**	**4**	**2**	**2**	**2**	**2**	**2**
		8/24/2005	0	126										
		8/25/2005	0	126	✓	✓			✓					
		8/26/2005	0	126	✓	✓			✓	✓	✓	✓	✓	✓
		8/27/2005	0	102	✓	✓	✓	✓	✓					
		8/28/2005	0	78	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		8/29/2005	0	54										
Humberto	**2**				**0**	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**1**	**1**
		9/12/2007	12	48										
		9/13/2007	0	36						✓	✓	✓	✓	✓
Ingrid	**4**				**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
		9/12/2007	12	126										
		9/13/2007	12	120										
		9/14/2007	12	96										
		9/15/2007	12	72										
Emily	**10**				**2**	**2**	**0**	**0**	**2**	**1**	**1**	**1**	**1**	**1**
		7/11/2005	0	126										
		7/12/2005	0	126										
		7/13/2005	0	126										
		7/14/2005	0	126										
		7/15/2005	0	126										
		7/16/2005	0	126										
		7/17/2005	0	108										
		7/18/2005	0	84										
		7/19/2005	0	60	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		7/20/2005	0	36	✓	✓	✓	✓	✓					
Ophelia	**11**				**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Storm	# of Cases	Forecast Date	Forecast Time	Hours w/ track	UWM1	UWM2	UWM3	UWM4	UWM5	PSU1	PSU2	PSU3	PSU4	PSU5
		9/6/2005	12	126										
		9/7/2005	12	126										
		9/8/2005	12	126										
		9/9/2005	12	126										
		9/10/2005	12	126										
		9/11/2005	12	126										
		9/12/2005	12	126										
		9/13/2005	12	126										
		9/14/2005	12	126										
		9/15/2005	12	126										
		9/16/2005	12	126										
Total	**69**				**17**	**17**	**2**	**2**	**17**	**9**	**9**	**9**	**9**	**9**

Appendix C: Inventories of tracked and evaluated forecasts. Check marks indicate cases for which the tracker produced a fix for all lead times that the storm was in its tropical phase. Those cases for which the tracker did not produce a fix for the entire tropical phase of the storm or the tracker output was cropped are indicated by numeral entries that correspond to the longest lead time for which the storm was verified. Those cases for which no lead times were included in the evaluation are indicated by blank cells.

Table C1: AOML inventory

Storm	# of Cases	Forecast Date	Forecast Time	Hours as TC	AOM1	AOM2	AOM5	AOM6
Wilma	**11**							
		10/16/2005	0	126	✓	✓	✓	✓
		10/17/2005	0	126	✓	110	✓	✓
		10/18/2005	0	126	✓	✓	✓	✓
		10/19/2005	0	126	✓	✓	✓	✓
		10/19/2005	12	126	✓	✓	✓	✓
		10/20/2005	0	126	113.5	118.5	✓	✓
		10/21/2005	0	114	✓	97	✓	✓
		10/22/2005	0	90	✓	80	✓	✓
		10/23/2005	0	66	✓	✓	✓	✓
		10/24/2005	0	42	✓	✓	✓	✓
		10/25/2005	0	18	13	✓	✓	✓
Philippe	**6**							
		9/17/2005	12	126	✓	✓	✓	✓
		9/18/2005	12	120	✓	✓	✓	✓
		9/19/2005	12	96	✓	✓	✓	✓
		9/20/2005	12	72	✓	✓	✓	✓
		9/21/2005	12	48	✓	✓	✓	✓
		9/22/2005	12	24	✓	✓	✓	✓
Felix	**8**							
		8/31/2007	12	114		✓	✓	✓
		9/1/2007	12	90	✓	14.5	✓	✓
		9/2/2007	0	78	✓	✓	✓	✓
		9/2/2007	6	72	✓	✓	✓	✓
		9/2/2007	12	66	✓	✓	✓	✓
		9/2/2007	18	60	✓	✓	✓	✓
		9/3/2007	0	54	✓	✓	✓	✓
		9/3/2007	12	42	✓	✓	✓	✓
Rita	**7**							
		9/18/2005	0	126			✓	✓
		9/19/2005	0	126	✓	✓	✓	✓
		9/20/2005	0	126	✓	✓	✓	✓
		9/21/2005	0	120	✓	✓	✓	✓
		9/22/2005	0	96	88	86	✓	✓
		9/23/2005	0	72	✓	✓	✓	✓
		9/24/2005	0	48	✓	✓	✓	✓
Karen	**4**							
		9/25/2007	0	102	✓	12	✓	✓

Storm	# of Cases	Forecast Date	Forecast Time	Hours as TC	AOM1	AOM2	AOM5	AOM6
		9/26/2007	0	78	✓	✓	✓	✓
		9/27/2007	0	64	✓	✓	✓	✓
		9/28/2007	0	30	✓	✓	✓	✓
Katrina	**6**							
		8/24/2005	0	126	✓	✓	✓	✓
		8/25/2005	0	126	✓	✓	✓	✓
		8/26/2005	0	114	✓	✓	✓	✓
		8/27/2005	0	90	✓	✓	✓	✓
		8/28/2005	0	66	✓	✓	✓	✓
		8/29/2005	0	42	✓	✓	✓	✓
Humberto	**2**							
		9/12/2007	12	36	✓	✓	✓	✓
		9/13/2007	0	24	✓	✓	✓	✓
Ingrid	**4**							
		9/12/2007	12	108	✓	✓	✓	✓
		9/13/2007	12	84	✓	✓	✓	✓
		9/14/2007	12	60	✓	✓	✓	✓
		9/15/2007	12	36	✓	✓	✓	✓
Emily	**10**							
		7/11/2005	0	126	18	19.5	✓	✓
		7/12/2005	0	126		✓	✓	✓
		7/13/2005	0	126	123	✓	✓	✓
		7/14/2005	0	126	28	✓	✓	✓
		7/15/2005	0	126	✓	✓	✓	✓
		7/16/2005	0	126	✓	114.5	✓	✓
		7/17/2005	0	108	96	97.5	✓	✓
		7/18/2005	0	84	54.5	74.5	✓	✓
		7/19/2005	0	60	46	44.5	✓	✓
		7/20/2005	0	36	27.5	24	✓	✓
Ophelia	**11**							
		9/6/2005	12	126	✓	✓	✓	✓
		9/7/2005	12	126	✓	✓	✓	✓
		9/8/2005	12	126	✓	✓	✓	✓
		9/9/2005	12	126	✓	✓	✓	✓
		9/10/2005	12	126	✓	✓	✓	✓
		9/11/2005	12	126	✓	✓	✓	✓
		9/12/2005	12	126	124	✓	✓	✓
		9/13/2005	12	102	72	✓	✓	✓
		9/14/2005	12	78	✓	✓	✓	✓
		9/15/2005	12	54	✓	✓	✓	✓
		9/16/2005	12	30	✓	✓	✓	✓

Table C2: MMM inventory

Storm	# of Cases	Forecast Date	Forecast Time	Hours as TC	MMM1	MMM3	MMM4
Wilma	**11**						
		10/16/2005	0	126	65.5		
		10/17/2005	0	126	✓	✓	0
		10/18/2005	0	126	✓	✓	✓
		10/19/2005	0	126	✓	✓	✓
		10/19/2005	12	126	✓	✓	✓
		10/20/2005	0	126	120	125	✓
		10/21/2005	0	114	108	108	108
		10/22/2005	0	90	79	79	62.5
		10/23/2005	0	66	90	58	90
		10/24/2005	0	42	39	36	66
		10/25/2005	0	18	16	16	16
Philippe	**6**						
		9/17/2005	12	126	✓		
		9/18/2005	12	120	✓	✓	✓
		9/19/2005	12	96	✓	35	✓
		9/20/2005	12	72	✓	✓	✓
		9/21/2005	12	48	✓	✓	21
		9/22/2005	12	24	✓	✓	✓
Felix	**8**						
		8/31/2007	12	114		15.5	15.5
		9/1/2007	12	90	23.5	✓	✓
		9/2/2007	0	78	✓	✓	✓
		9/2/2007	6	72	✓	✓	✓
		9/2/2007	12	66	✓	✓	✓
		9/2/2007	18	60	✓	✓	✓
		9/3/2007	0	54	✓	✓	✓
		9/3/2007	12	42	✓	✓	✓
Rita	**7**						
		9/18/2005	0	126	✓	✓	✓
		9/19/2005	0	126	✓	✓	✓
		9/20/2005	0	126	✓		
		9/21/2005	0	120	✓	✓	✓
		9/22/2005	0	96	✓	✓	✓
		9/23/2005	0	72	✓	✓	✓
		9/24/2005	0	48	✓	✓	✓
Karen	**4**						
		9/25/2007	0	102			
		9/26/2007	0	78	✓	✓	✓
		9/27/2007	0	64	✓	✓	✓
		9/28/2007	0	30	✓	20	36.5
Katrina	**6**						
		8/24/2005	0	126	✓	✓	✓
		8/25/2005	0	126	✓	✓	✓
		8/26/2005	0	114	✓	✓	✓

Storm	# of Cases	Forecast Date	Forecast Time	Hours as TC	MMM1	MMM3	MMM4
		8/27/2005	0	90	✓	✓	✓
		8/28/2005	0	66	✓	✓	✓
		8/29/2005	0	42	✓	✓	✓
Humberto	**2**						
		9/12/2007	12	36	✓	✓	✓
		9/13/2007	0	24	✓	✓	✓
Ingrid	**4**						
		9/12/2007	12	108	✓	✓	✓
		9/13/2007	12	84	✓	✓	✓
		9/14/2007	12	60	✓	✓	✓
		9/15/2007	12	36	✓	✓	✓
Emily	**10**						
		7/11/2005	0	126	✓	✓	2
		7/12/2005	0	126	✓	✓	✓
		7/13/2005	0	126	✓	✓	✓
		7/14/2005	0	126		✓	✓
		7/15/2005	0	126	✓	✓	✓
		7/16/2005	0	126	115.5	99	99
		7/17/2005	0	108	91	75	75
		7/18/2005	0	84	67	51	55
		7/19/2005	0	60	47.5	33	33
		7/20/2005	0	36	24.5	9	9
Ophelia	**11**						
		9/6/2005	12	126	✓	46	8.5
		9/7/2005	12	126	✓	✓	✓
		9/8/2005	12	126	✓	✓	✓
		9/9/2005	12	126	✓	✓	✓
		9/10/2005	12	126	✓	✓	✓
		9/11/2005	12	126	✓	✓	✓
		9/12/2005	12	126	✓	✓	✓
		9/13/2005	12	102	88.5	79	79
		9/14/2005	12	78	✓	✓	✓
		9/15/2005	12	54	✓	✓	✓
		9/16/2005	12	30	✓	✓	✓

Table C3: NRL inventory

Storm	# Cases	Forecast Date	Forecast	Hours as TC	NRL1	NRL2	NRL5
Wilma	**11**						
		10/16/2005	0	126	✓	16.5	✓
		10/17/2005	0	126	✓	✓	✓
		10/18/2005	0	126	✓	✓	✓
		10/19/2005	0	126	✓	✓	✓
		10/19/2005	12	126	✓	✓	✓
		10/20/2005	0	126	✓	✓	✓
		10/21/2005	0	114	111	110	109.5
		10/22/2005	0	90	✓	87.5	✓
		10/23/2005	0	66	✓	✓	✓
		10/24/2005	0	42	✓	35	✓
		10/25/2005	0	18	17	15.5	17.5
Philippe	**6**						
		9/17/2005	12	126	✓	✓	✓
		9/18/2005	12	120	✓	✓	✓
		9/19/2005	12	96	✓	✓	✓
		9/20/2005	12	72	✓	✓	✓
		9/21/2005	12	48	✓	0.5	0
		9/22/2005	12	24	7	✓	✓
Felix	**8**						
		8/31/2007	12	114			
		9/1/2007	12	90			
		9/2/2007	0	78	✓	✓	✓
		9/2/2007	6	72	5.5	5.5	5.5
		9/2/2007	12	66	✓	✓	✓
		9/2/2007	18	60	19	✓	✓
		9/3/2007	0	54	✓	✓	✓
		9/3/2007	12	42	✓	✓	✓
Rita	**7**						
		9/18/2005	0	126			
		9/19/2005	0	126		49	49
		9/20/2005	0	126	✓	✓	✓
		9/21/2005	0	120	✓	✓	✓
		9/22/2005	0	96	✓	✓	✓
		9/23/2005	0	72	✓	42.5	57
		9/24/2005	0	48	✓	47	✓
Karen	**4**						
		9/25/2007	0	102	8	0.5	5
		9/26/2007	0	78	✓	✓	✓
		9/27/2007	0	64	✓	✓	✓
		9/28/2007	0	30	✓	✓	✓
Katrina	**6**						
		8/24/2005	0	126	14	93	ü
		8/25/2005	0	126	✓	✓	✓
		8/26/2005	0	114	✓	✓	✓
		8/27/2005	0	90	✓	✓	✓
		8/28/2005	0	66	✓	✓	✓

		8/29/2005	0	42	✓	✓	✓
Humberto	**2**						
		9/12/2007	12	36	26.5	24.5	9
		9/13/2007	0	24	✓	✓	✓
Ingrid	**4**						
		9/12/2007	12	108	43	✓	✓
		9/13/2007	12	84	18.5	✓	✓
		9/14/2007	12	60	✓	✓	✓
		9/15/2007	12	36	✓	✓	✓
Emily	**10**						
		7/11/2005	0	126			
		7/12/2005	0	126	✓	9.5	23
		7/13/2005	0	126	✓	19.5	53.5
		7/14/2005	0	126	✓	✓	✓
		7/15/2005	0	126	0	7.5	0
		7/16/2005	0	126	✓	✓	✓
		7/17/2005	0	108	✓	✓	✓
		7/18/2005	0	84	✓	✓	✓
		7/19/2005	0	60	✓	✓	✓
		7/20/2005	0	36	✓	✓	✓
Ophelia	**11**						
		9/6/2005	12	126	24	11	7
		9/7/2005	12	126	✓	✓	✓
		9/8/2005	12	126	✓	✓	✓
		9/9/2005	12	126	✓	✓	✓
		9/10/2005	12	126	✓	125.5	✓
		9/11/2005	12	126	✓	✓	✓
		9/12/2005	12	126	✓	117.5	119
		9/13/2005	12	102	✓	86.5	87
		9/14/2005	12	78	✓	✓	✓
		9/15/2005	12	54			
		9/16/2005	12	30			

Table C4: PSU inventory

Storm	# Cases	Forecast Date	Forecast	Hours as TC	PSU1	PSU2	PSU3	PSU4	PSU5
Wilma	**11**								
		10/16/2005	0	126					
		10/17/2005	0	126					
		10/18/2005	0	126					
		10/19/2005	0	126					
		10/19/2005	12	126					
		10/20/2005	0	126					
		10/21/2005	0	114	✓	✓	✓	✓	✓
		10/22/2005	0	90					
		10/23/2005	0	66	✓	59	63.5	60	✓
		10/24/2005	0	42					
		10/25/2005	0	18					
Philippe	**6**								
		9/17/2005	12	126					
		9/18/2005	12	120					
		9/19/2005	12	96					
		9/20/2005	12	72					
		9/21/2005	12	48					
		9/22/2005	12	24					
Felix	**8**								
		8/31/2007	12	114					
		9/1/2007	12	90					
		9/2/2007	0	78					
		9/2/2007	6	72					
		9/2/2007	12	66					
		9/2/2007	18	60					
		9/3/2007	0	54					
		9/3/2007	12	42					
Rita	**7**								
		9/18/2005	0	126					
		9/19/2005	0	126					
		9/20/2005	0	126	✓	✓	✓	✓	✓
		9/21/2005	0	120	✓	✓	✓	✓	✓
		9/22/2005	0	96	✓	✓	✓	✓	✓
		9/23/2005	0	72					
		9/24/2005	0	48					
Karen	**4**								
		9/25/2007	0	102					
		9/26/2007	0	78					
		9/27/2007	0	64					
		9/28/2007	0	30					
Katrina	**6**								
		8/24/2005	0	126					
		8/25/2005	0	126					
		8/26/2005	0	114	✓	✓	✓	✓	✓
		8/27/2005	0	90					
		8/28/2005	0	66	✓	✓	✓	✓	✓

		8/29/2005	0	42					
Humberto	**2**								
		9/12/2007	12	36					
		9/13/2007	0	24	✓	✓	✓	✓	✓
Ingrid	**4**								
		9/12/2007	12	108					
		9/13/2007	12	84					
		9/14/2007	12	60					
		9/15/2007	12	36					
Emily	**10**								
		7/11/2005	0	126					
		7/12/2005	0	126					
		7/13/2005	0	126					
		7/14/2005	0	126					
		7/15/2005	0	126					
		7/16/2005	0	126					
		7/17/2005	0	108					
		7/18/2005	0	84					
		7/19/2005	0	60	51	50.5	48	✓	✓
		7/20/2005	0	36					
Ophelia	**11**								
		9/6/2005	12	126					
		9/7/2005	12	126					
		9/8/2005	12	126					
		9/9/2005	12	126					
		9/10/2005	12	126					
		9/11/2005	12	126					
		9/12/2005	12	126					
		9/13/2005	12	102					
		9/14/2005	12	78					
		9/15/2005	12	54					
		9/16/2005	12	30					

Table C5: URI inventory

Storm	# Cases	Forecast Date	Forecast Time	Hours as TC	URI1	URI2
Wilma	**11**					
		10/16/2005	0	126	✓	✓
		10/17/2005	0	126	✓	✓
		10/18/2005	0	126	✓	✓
		10/19/2005	0	126	✓	124.5
		10/19/2005	12	126	105.5	✓
		10/20/2005	0	126	✓	✓
		10/21/2005	0	114	✓	✓
		10/22/2005	0	90	✓	✓
		10/23/2005	0	66	✓	✓
		10/24/2005	0	42	✓	✓
		10/25/2005	0	18	✓	✓
Philippe	**6**					
		9/17/2005	12	126	35	✓
		9/18/2005	12	120	✓	✓
		9/19/2005	12	96	✓	✓
		9/20/2005	12	72	✓	✓
		9/21/2005	12	48	✓	✓
		9/22/2005	12	24	✓	✓
Felix	**8**					
		8/31/2007	12	114	✓	✓
		9/1/2007	12	90	14.5	✓
		9/2/2007	0	78	✓	✓
		9/2/2007	6	72	✓	✓
		9/2/2007	12	66	✓	✓
		9/2/2007	18	60	✓	✓
		9/3/2007	0	54	✓	✓
		9/3/2007	12	42	✓	✓
Rita	**7**					
		9/18/2005	0	126	✓	1.5
		9/19/2005	0	126	✓	✓
		9/20/2005	0	126	✓	✓
		9/21/2005	0	120	✓	✓
		9/22/2005	0	96	✓	✓
		9/23/2005	0	72	✓	✓
		9/24/2005	0	48	✓	✓
Karen	**4**					
		9/25/2007	0	102	✓	✓
		9/26/2007	0	78	✓	✓
		9/27/2007	0	64	✓	✓
		9/28/2007	0	30	✓	✓
Katrina	**6**					
		8/24/2005	0	126	✓	✓
		8/25/2005	0	126	✓	✓

		8/26/2005	0	114	✓	✓
		8/27/2005	0	90	✓	✓
		8/28/2005	0	66	✓	✓
		8/29/2005	0	42	✓	✓
Humberto	**2**					
		9/12/2007	12	36	✓	✓
		9/13/2007	0	24	✓	✓
Ingrid	**4**					
		9/12/2007	12	108	✓	✓
		9/13/2007	12	84	✓	✓
		9/14/2007	12	60	44.5	38.5
		9/15/2007	12	36	23	23
Emily	**10**					
		7/11/2005	0	126	33	59.5
		7/12/2005	0	126	30.5	64
		7/13/2005	0	126	✓	✓
		7/14/2005	0	126	✓	88.5
		7/15/2005	0	126	✓	✓
		7/16/2005	0	126	✓	✓
		7/17/2005	0	108	✓	✓
		7/18/2005	0	84	✓	✓
		7/19/2005	0	60	✓	✓
		7/20/2005	0	36	0	0
Ophelia	**11**					
		9/6/2005	12	126	✓	✓
		9/7/2005	12	126	✓	✓
		9/8/2005	12	126	✓	✓
		9/9/2005	12	126	✓	✓
		9/10/2005	12	126	✓	✓
		9/11/2005	12	126	✓	✓
		9/12/2005	12	126	✓	✓
		9/13/2005	12	102	✓	✓
		9/14/2005	12	78	✓	✓
		9/15/2005	12	54	✓	✓
		9/16/2005	12	30	✓	✓

Table C6: UWM inventory

Storm	# of Cases	Forecast Date	Forecast Time	Hours as TC	UWM1	UMW2	UMW3	UWM4	UMW5
Wilma	**11**								
		10/16/2005	0	126	✓	✓			✓
		10/17/2005	0	126	✓	✓			✓
		10/18/2005	0	126	✓	✓			✓
		10/19/2005	0	126	✓	✓			✓
		10/19/2005	12	126	✓	✓			✓
		10/20/2005	0	126					
		10/21/2005	0	114					
		10/22/2005	0	90					
		10/23/2005	0	66					
		10/24/2005	0	42					
		10/25/2005	0	18	✓	✓			✓
Philippe	**6**								
		9/17/2005	12	126					
		9/18/2005	12	120					
		9/19/2005	12	96					
		9/20/2005	12	72					
		9/21/2005	12	48	✓	✓			✓
		9/22/2005	12	24	✓	✓			✓
Felix	**8**								
		8/31/2007	12	114					
		9/1/2007	12	90					
		9/2/2007	0	78					
		9/2/2007	6	72					
		9/2/2007	12	66					
		9/2/2007	18	60					
		9/3/2007	0	54					
		9/3/2007	12	42					
Rita	**7**								
		9/18/2005	0	126					
		9/19/2005	0	126	✓	✓			✓
		9/20/2005	0	126	✓	✓			✓
		9/21/2005	0	120	✓	✓			✓
		9/22/2005	0	96					
		9/23/2005	0	72					
		9/24/2005	0	48					
Karen	**4**								
		9/25/2007	0	102					
		9/26/2007	0	78					
		9/27/2007	0	64					
		9/28/2007	0	30					
Katrina	**6**								
		8/24/2005	0	126					
		8/25/2005	0	126	✓	✓			✓
		8/26/2005	0	114	✓	✓			✓

Storm	# of Cases	Forecast Date	Forecast Time	Hours as TC	UWM1	UMW2	UMW3	UWM4	UMW5
		8/27/2005	0	90	✓	✓	✓	✓	✓
		8/28/2005	0	66	✓	✓	✓	✓	✓
		8/29/2005	0	42					
Humberto	**2**								
		9/12/2007	12	36					
		9/13/2007	0	24					
Ingrid	**4**								
		9/12/2007	12	108					
		9/13/2007	12	84					
		9/14/2007	12	60					
		9/15/2007	12	36					
Emily	**10**								
		7/11/2005	0	126					
		7/12/2005	0	126					
		7/13/2005	0	126					
		7/14/2005	0	126					
		7/15/2005	0	126					
		7/16/2005	0	126					
		7/17/2005	0	108					
		7/18/2005	0	84					
		7/19/2005	0	60	36	36			36
		7/20/2005	0	36	12	12			12
Ophelia	**11**								
		9/6/2005	12	126					
		9/7/2005	12	126					
		9/8/2005	12	126					
		9/9/2005	12	126					
		9/10/2005	12	126					
		9/11/2005	12	126					
		9/12/2005	12	126					
		9/13/2005	12	102					
		9/14/2005	12	78					
		9/15/2005	12	54					
		9/16/2005	12	30					

Appendix D: Inventory of statistically significant differences for verification over land and water. Table entries are defined as: H = SS difference for which the errors associated with the high-resolution configuration are smaller than that for the low-resolution configuration, L = SS difference for which the errors associated with the low-resolution configuration are smaller than that for the high-resolution configuration, dash = no SS difference for a sample size greater than 11, and blank = sample size less than 11.

Table D1: AOML

AOM (land and water): Median for pairwise difference with 95% CIs																					
Forecast hour		0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Track		-	-	-	-	-	H	H	H	H	-	-	-	-	-	-	-	-	-	-	-	-
Along Track		-	-	-	-	-	H	-	H	-	-	-	-	-	-	-	-	-	-	-	-	-
Cross Track		-	-	-	-	-	-	H	-	H	-	-	-	-	-	-	-	-	-	-	-	-
Intensity		H	H	-	-	H	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Wind Radii	NE34	-	-	-	-	-	-	-	-	-	-	-	-	-	L	L	L	-	-	-	-	L
	NE50	-	-	-	-	-	-	-	L	-	L	-	L	L	L	-	L	-	-	-	-	-
	NE64	-	-	-	-	-	-	-	L	-	-	-	-	L	L	L						
	SE34	-	-	-	-	-	-	-	-	L	L	L	L	L	-	L	L	L	L	-	L	L
	SE50	-	-	-	-	-	-	-	-	-	L	L	L	L	L	-	L	-	-	L	-	-
	SE64	-	-	-	-	-	-	-	-	-	-	-	-	-	L	L						
	SW34	-	-	-	-	-	-	-	-	-	L	-	L	-	L	L	-	L	L	L	-	L
	SW50	-	-	-	-	-	-	-	-	-	-	L	L	L	-	L	L	-	L	L	-	-
	SW64	-	-	H	-	-	H	-	-	-	-	-	L	-	L	L						
	NW34	-	-	-	-	-	L	-	-	-	-	L	L	-	-	-	-	-	-	-	L	-
	NW50	-	-	-	-	-	-	-	L	L	-	L	L	L	-	-	-	-	-	-	-	L
	NW64	-	-	-	-	-	L	-	L	L	-	-	L	-	L	L						

Table D2: MMM

MMM(land and water): Median for pairwise difference with 95% CIs																				
Forecast hour	0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Track	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	H	H	H	H	-
Along Track	-	-	-	-	-	L	-	-	-	-	-	-	-	-	-	-	H	-	-	-	-
Cross Track	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	H	-	-	H	-
Intensity	-	-	-	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table D3: NRL

NRL (land and water): Median for pairwise difference with 95% CIs																					
Forecast		0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Track		-	-	-	-	H	-	-	L	-	L	-	-	-	-	-	-	L	-	-	-	-
Along Track		-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	-	L	-	-	L	-
Cross Track		-	-	-	-	-	-	-	-	-	-	-	-	L	-	-	-	-	-	-	-	H
Intensity		H	H	-	-	H	-	-	H	-	-	-	-	-	-	-	-	-	-	-	-	-
Wind Radii	NE34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	NE50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
	NE64	-	L	L	-	-	-	L	L	-	-	-	L	-		-	L					
	SE34	-	-	-	-	-	-	-	-	-	L	-	-	-	L	-	L	-	-	-	L	-
	SE50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	-	-				
	SE64	-	-	-	-	-	-	-	-	-	-	-	-	-		-						
	SW34	-	-	-	L	-	L	-	-	-	-	L	-	-	-	-	-	-	-	-	-	-
	SW50	-	-	-	-	-	-	-	-	-	H	-	-	-	H	-	-	-				
	SW64	-	-	-	-	-	-	-	-	-	-	-	-	-		-						
	NW34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	NW50	-	-	-	L	-	-	-	-	-	-	-	-	-	-	-	-	-				
	NW64	-	-	-	-	-	-	-	-	-	-	-	-	-		-						

Table D4: URI

URI (land and water): Median for pairwise difference with 95% CIs																					
Forecast hour		0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Track		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Along Track		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cross Track		-	-	-	-	-	-	-	-	H	-	-	-	H	-	-	-	-	-	-	-	-
Intensity		H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	L	-	-	-	-
Wind Radii	NE34	-	L	-	L	-	-	L	L	-	-	-	L	-	L	-	L	-	L	L	-	L
	NE50	H	-	-	-	L	-	-	-	-	-	L	L	-	-	-	L	-	-	-	-	-
	NE64	-	-	-	-	-	L	L	L	-	L	L	-	-	L	L	-	-	-	-	-	-
	SE34	H	L	L	L	-	L	-	-	-	-	L	L	L	L	L	-	L	L	L	L	L
	SE50	-	-	-	L	-	-	-	-	-	-	L	L	L	L	L	L	L	-	-	-	L
	SE64	-	-	-	-	L	L	L	L	-	-	L	L	-	L	L	L	-	-	-	-	-
	SW34	-	-	-	-	-	-	-	-	-	-	-	-	L	-	L	L	-	L	L	-	-
	SW50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	L	L	L	L	-	-	-
	SW64	-	-	-	-	-	-	-	L	-	-	-	-	-	-	L	-	-	-	-	-	-
	NW34	-	-	L	-	L	L	-	L	-	L	L	-	L	-	L	-	-	L	-	-	-
	NW50	-	-	-	L	L	-	-	L	L	L	L	L	L	-	-	L	-	-	-	-	-
	NW64	-	-	-	L	-	L	-	L	-	L	-	L	-	-	-	-	-	-	-	-	-

Table D5: UWM

UWM (land and water): Median for pairwise difference with 95% CIs																					
Forecast hour		0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120
Track		-	-	-	-	-	-	-	-	-	-	-	-	-	L	-	-					
Along Track		-	-	-	-	-	-	-	-	-	-	L	-	-	-	-	-					
Cross Track		-	-	-	-	-	-	-	-	-	-	-	-	-	L	-	-					
Intensity		H	-	-	-	-	-	-	-	H	-	H	-	H	H	H	-					
Wind Radii	NE34	-	-	-	-	-	-	-	-	-	-	L		-	-							
	NE50		-	-			-	-	-		L	-	-									
	NE64								-				-									
	SE34	-	-	-	-	L	-	-	-	-	L	L	L	-	-							
	SE50		-	-			-	-	L		-	-	-									
	SE64								-				-									
	SW34	-	-	-	-	-	-	-	-	-	-	-	L	L	L							
	SW50		-	-			-	-	-		-	-	L									
	SW64								L				L									
	NW34	-	-	-	-	L	-	-	-	-	-	L	-	-	-							
	NW50		H	-			-	-	-		-	-	-									
	NW64								-				-									

Appendix E: List of acronyms

ACARS - Aircraft Communications Addressing and Reporting System
AHW – Advanced Hurricane WRF
ARW – Advanced Research WRF
AOML - Atlantic Oceanographic and Meteorological Laboratory
ATCF – Automated Tropical Cyclone Forecast System
CIRA – Cooperative Institute for Research in the Atmosphere
COAMPS - Coupled Ocean/Atmosphere Mesoscale Prediction System
COAMPS-TC - Coupled Ocean/Atmosphere Mesoscale Prediction System – Tropical Cyclone
CI – Confidence Interval
CSI – Critical Success Index
DTC – Developmental Testbed Center
DOD – Department of Defense
EMC – Environmental Modeling Center
EnKF – Ensemble Kalman Filter
ESRL – Earth System Research Laboratory
ESSL – Earth and Sun Systems Laboratory
FAR – False Alarm Rate
GFDL – Geophysical Fluid Dynamics Laboratory
GFS – Global Forecasting System
GSFC – Goddard Space Flight Center
GRIB1 – Gridded Binary 1
HFIP - Hurricane Forecast Improvement Project
HRH – High Resolution Hurricane
HWRF – Hurricane WRF
HWRF-X – Experimental Hurricane WRF
HYCOM – Hybrid Coordinate Ocean Model
JTWC – Joint Typhoon Warning Center
MAE – Mean absolute error
ME – Mean error
MCSST – Multi-Channel Sea Surface Temperature
MMM – Mesoscale and Microscale Model
MRF – Medium-Range Forecast
MSLP – Mean Sea Level Pressure
MSSW – Maximum Sustained Surface Wind
MSW – Maximum Surface Wind
NASA – National Aeronautics and Space Administration
NAVDAS – NRL Atmospheric Variational Data Assimilation System
NCAR – National Center for Atmospheric Research
NCL – NCAR Command Language
NCODA – NRL Coupled Ocean Data Assimilation
NCEP – National Centers for Environmental Prediction
NE - Northeast
NHC – National Hurricane Center
NMM – Nonhydrostatic Mesoscale Model
NRL – National Research Laboratory

NOAA – National Oceanic and Atmospheric Administration
NOGAPS – Navy Operational Global Atmospheric Prediction System
NW - Northwest
NWP – Numerical Weather Prediction
NWS – National Weather Service
OCD5 – Decay-SHIFOR5
PC – Proportion Correct
POD – Probability of detection
POM – Princeton Ocean Model
PSU – Pennsylvania State University
RAINEX - Hurricane Rainband and Intensity Change Experiment
RMSE – Root-Mean Squared Error
RI – Rapid Intensification
RRTM – Rapid Radiative Transfer Model
RSMAS – Rosenstiel School of Marine and Atmospheric Science
RW – Rapid Weakening
SAIC – Science Applications International Corporation
SE - Southeast
SS – Statistically Significant
SSM/I – Special Sensor Microwave Imager
SST – Sea Surface Temperature
SW - Southwest
TC – Tropical Cyclone
TKE – Turbulent Kinetic Energy
URI – University of Rhode Island
UWM – University of Wisconsin-Madison
UW-NMS – University of Wisconsin Non-hydrostatic Modeling System
WPS – WRF Pre-processing System
WRF – Weather Research and Forecasting
WRF-Var – WRF Variational Data Assimilation System
WSM – WRF Single-Moment
YSU - Yonsei University