Toward Improving High-Resolution Numerical Hurricane Forecasting: Influence of Model Horizontal Grid Resolution, Initialization, and Physics

SUNDARARAMAN G. GOPALAKRISHNAN,* STANLEY GOLDENBERG,* THIAGO QUIRINO,* XUEJIN ZHANG,+ FRANK MARKS JR.,* KAO-SAN YEH,+ ROBERT ATLAS,# AND VIJAY TALLAPRAGADA@

* *NOAA/Atlantic Oceanographic and Meteorological Laboratory/Hurricane Research Division, Miami, Florida*
+ *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida*
# *NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*
@ *NOAA/National Centers for Environmental Prediction/Environmental Modeling Center, Washington, D.C.*

(Manuscript received 16 May 2011, in final form 28 November 2011)

ABSTRACT

This paper provides an account of the performance of an experimental version of the Hurricane Weather Research and Forecasting system (HWRFX) for 87 cases of Atlantic tropical cyclones during the 2005, 2007, and 2009 hurricane seasons. The HWRFX system was used to study the influence of model grid resolution, initial conditions, and physics. For each case, the model was run to produce 126 h of forecast with two versions of horizontal resolution, namely, (i) a parent domain at a resolution of about 27 km with a 9-km moving nest (27:9) and (ii) a parent domain at a resolution of 9 km with a 3-km moving nest (9:3). The former was selected to be consistent with the current operational resolution, while the latter is the first step in testing the impact of finer resolutions for future versions of the operational model. The two configurations were run with initial conditions for tropical cyclones obtained from the operational Geophysical Fluid Dynamics Laboratory (GFDL) and HWRF models. Sensitivity experiments were also conducted with the physical parameterization scheme. The study shows that the 9:3 HWRFX system using the GFDL initial conditions and a system of physics similar to the operational version (HWRF) provides the best results in terms of both track and intensity prediction. Use of the HWRF initial conditions in the HWRFX model provides reasonable skill, particularly when used in cases with initially strong storms (hurricane strength). However, initially weak storms (below hurricane strength) posed special challenges for the models. For the weaker storm cases, none of the predictions from the HWRFX runs or the operational GFDL forecasts provided any consistent improvement when compared to the operational Statistical Hurricane Intensity Prediction Scheme with an inland decay component (DSHIPS).

1. Introduction

The past decades have been marked by significant advances in operational weather prediction models such as the National Oceanic and Atmospheric Administration's (NOAA) Global Forecast System (GFS), the NOAA/Geophysical Fluid Dynamics Laboratory's (GFDL) regional hurricane forecasting system, the U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS), the European Centre for Medium-Range Weather Forecasts (ECMWF) model, and the Met Office model (UKMET). Such advances have contributed greatly to a steady improvement in the official tropical cyclone (TC) track forecasts issued by the NOAA/National Weather Service's (NWS) National Hurricane Center (NHC), resulting in a substantial reduction in track forecast errors. This, in turn, has reduced warning and evacuation areas, thereby saving lives and resources (Rappaport et al. 2009).

Forecasting intensity changes is also extremely important, especially in the case of storms that rapidly intensify or weaken just prior to landfall (e.g., TCs Charley, 2004; Katrina and Wilma, 2005; Humberto, 2007; Karl, 2010). However, forecasting intensity changes in TCs is a complex and challenging multiscale problem. While cloud-resolving numerical models using a horizontal grid resolution of 1–3 km are starting to show some skill in predicting intensity changes in individual cases (e.g., Chen et al. 2011), it is not clear at this time if such high-resolution models can provide intensity guidance with

Corresponding author address: S. G. Gopalakrishnan, AOML/HRD, 4301 Rickenbacker Cswy., Miami, FL 33149.
E-mail: Sundararaman.G.Gopalakrishnan@noaa.gov

DOI: 10.1175/WAF-D-11-00055.1

fidelity. The goal is to produce consistently reliable results in the real-time (operational) venue. Lack of skill in numerical intensity forecasting is often associated with inaccurate initial conditions or limitations in modeling the physical processes within and around hurricanes. There is also debate as to what constitutes a reasonable horizontal resolution for use in forecasting hurricane intensity changes on a day-to-day basis (Gopalakrishnan et al. 2011).

The Weather Research and Forecast (WRF) system for hurricane prediction (HWRF) became operational at the National Centers for Environmental Prediction (NCEP) in 2007. This advanced hurricane prediction system was developed at the NWS/Environmental Modeling Center (EMC) to address the nation's next-generation hurricane forecast problems. In this study, an experimental version of HWRF (HWRFX) is used to explore three important factors (i.e., model grid resolution, initial conditions, and model physics) that may influence the accuracy of track and intensity forecasts and provide guidance for improvements to the operational HWRF system.

2. Background

The GFDL regional hurricane prediction system originated as a research model in the 1970s (Kurihara and Tuleya 1974; Kurihara 1975; Tuleya and Kurihara 1975, 1978, 1982). In the mid-1980s, the Hurricane Dynamics Group at GFDL began a 10-yr effort to transform their research hurricane model into an operational hurricane forecasting tool for NWS. As part of that effort, several years were spent in developing a technique to insert a more realistic and model-consistent vortex into the global analysis. The initialization of hurricanes in the GFDL model, which is unique to this model and is relevant to the discussion later in this study, uses a vortex replacement strategy that consists of three major steps: 1) interpolate the global analysis fields from GFS onto the operational GFDL hurricane model grid, 2) remove the GFS vortex from the global analysis, and 3) add a high-resolution, model-consistent vortex (Kurihara et al. 1995). Since 1995, the GFDL hydrostatic hurricane prediction system has been used operationally by NHC and is regarded by NHC as one of the most skillful and reliable dynamical models for track forecasts. Since 2005, the model has demonstrated improved skill in intensity predictions, especially after the atmosphere–ocean coupled system (Ginis et al. 1999) and increases in vertical and horizontal resolution were implemented (Bender et al. 2007). A thorough summary of the modeling system and its performance is provided in Bender et al. (2007).

WRF is a general purpose, multi-institutional mesoscale modeling system. NCEP's Nonhydrostatic Mesoscale Model (NMM; Janjić et al. 2001; Janjić 2003) is a dynamical core option within the WRF model initiative. A moving-nest capability was created within the WRF-NMM system to address the hurricane problem (Gopalakrishnan et al. 2006). This dynamical capability is the backbone of the HWRF system. The high-resolution nest is capable of capturing nonhydrostatic scales of motion within the hurricane inner core and has the potential to provide improved intensity guidance. Another advancement of the HWRF system over the GFDL model is the option to use the previous forecast cycle of the model vortex (when available) as the first guess and to assimilate inner-core observations. The GFDL model uses a basic mass-consistent bogus vortex whereas HWRF uses a model-consistent vortex from the previous cycle that has been relocated and adjusted toward current pressure and wind observations. This capability was introduced to provide a more realistic initial three-dimensional structure and is a critical ingredient toward advancing TC intensity/structure prediction (Liu et al. 2006). To build on the success of the GFDL model, the physics in the initial implementation of HWRF emulated the latest version of the GFDL system (Bender et al. 2007). In addition, the HWRF system, as with the GFDL model, is coupled to a three-dimensional version of the Princeton Ocean Model (POM) modified for hurricane applications over the Atlantic basin. The HWRF system, with 27- and 9-km parent and movable mesh grid resolutions (27:9; Fig. 1a), respectively, has been used as an operational hurricane forecasting tool at NWS since 2007. Documentation for the HWRF system (Gopalakrishnan et al. 2010) is available at the Development Test Bed Center (DTC), National Center for Atmospheric Research (NCAR), in Boulder, Colorado (http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRF_final_2-2_cm.pdf).

In addition to operational hurricane models, cloud-resolving models have more recently been used as research tools to help understand the hurricane intensity prediction problem. For example, in a series of explicit simulations of Hurricane Andrew (1992) using the fifth-generation Mesoscale Model (MM5), developed by the Pennsylvania State University (PSU) and NCAR at a resolution of about 6 km, several interesting features of the modeled hurricane (Liu et al. 1997, 1999; Yau et al. 2004; Zhang et al. 1999, 2000, 2001, 2002) compared well to the observed structure of hurricanes (Marks and Houze 1987; Marks et al. 1992; Willoughby 1979, 1990b,a). As verified against various observations and the poststorm-derived ''best track'' data (positions and intensities) provided by NHC, the PSU–NCAR MM5 simulations reasonably captured many of the features of the inner-core structure of the storm. In particular, the track, explosive deepening rate, minimum surface pressure preceding landfall, and strong surface winds near the shoreline

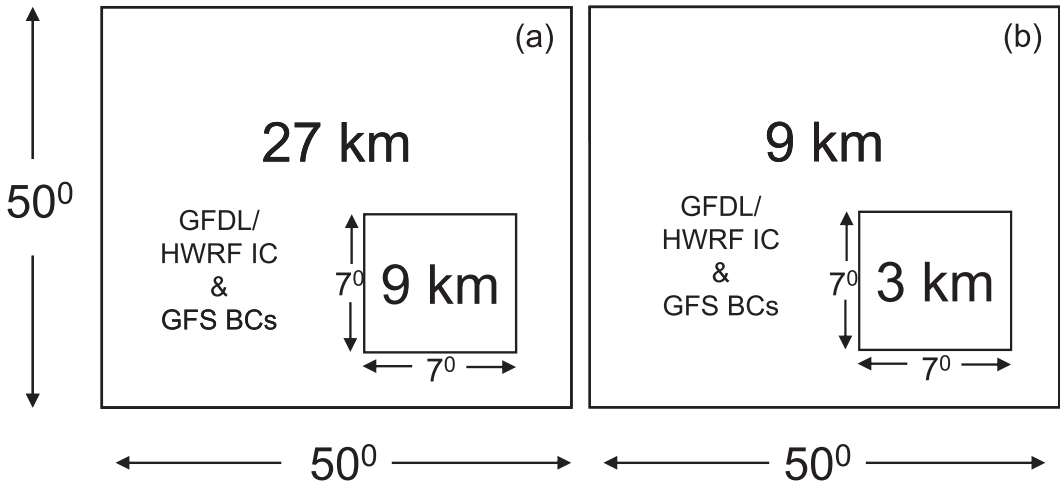

FIG. 1. Versions of the two domain configurations used in this study: parent domain at a resolution of (a) 27 km with a moving nest at 9 km and (b) 9 km with a moving nest at 3 km. The approximate sizes of the domains are shown as they appear in a rotated latitude–longitude grid system (i.e., without any projection to the actual latitude–longitude grid).

were all well reproduced by these simulations (Yau et al. 2004).

Other recent research suggests that hurricane intensity forecasts could be improved with very high model resolution (grid spacing ≤ 1 km in the horizontal) that adequately simulates hurricane inner-core structures such as the eyewall and rainbands (Davis et al. 2008, 2011; Rotunno et al. 2009). For example, in an effort to demonstrate the value of a high-resolution hurricane forecast, Davis et al. (2011) used the Advanced Hurricane WRF (AHW) model with two sets of 69 simulations (covering 10 Atlantic tropical cyclones) with horizontal resolutions of 12 and 1.33 km. Their statistically significant results indicated that higher resolution improved forecasts of both intensity and the hurricane's structural aspects. Using the results of Davis et al. (2011) as the basis, the authors used the AHW model to provide real-time forecasts during the 2009 season. The results were storm specific. While higher resolution improved forecasts of track and intensity, especially for some of the weaker storms like Erica and Danny, the impacts were mixed for other storms. These results raise the question of what controls the overall intensity and track predictions: resolution alone, physics, initial conditions, or all of these factors?

While operational track forecasts show substantial skill (on the order of 30%–40% at 72 h) versus the Climatology and Persistence (CLIPER) model utilized by NHC (e.g., Gross 1999), the skill in predicting intensity change using a dynamical model has been marginal (DeMaria and Kaplan 1999; Kaplan et al. 2010). To address this lack in intensity prediction skill, with an emphasis on rapid intensification (RI) events, NOAA established the Hurricane Forecast Improvement Project (HFIP) in 2007 (available online at http://www.nrc.noaa.gov/plans_docs/HFIP_Plan_073108.pdf).

HWRFX is a version of the HWRF system specifically adopted and developed at the Hurricane Research Division (HRD) of the Atlantic Oceanographic and Meteorological Laboratory (AOML) of NOAA to study the intensity change problem at cloud-resolving scales (about 1–3 km). This modeling system is supported by HFIP and complements the operational HWRF system. The data acquisition component, designed with Java and shell scripts, runs automatically and in parallel with the triggering mechanism. It continuously retrieves meteorological datasets required to perform the HWRFX forecasts. These datasets include the initial conditions and the 126-h forecast data of the operational Global Forecast System (GFS), the initial conditions of the operational GFDL hurricane model, and the initial conditions of the operational HWRF model. The HWRFX model can be run with at least two suites of physics options. The first suite mimics the operational HWRF suite of physics (Gopalakrishnan et al. 2010), and the second is a suite of physics that was developed as part of the high-resolution modeling effort at NCAR. The physics options used in this work are discussed in detail later in the text. As in the case of the GFDL model developments, HWRFX can be run in both an idealized research framework (Gopalakrishnan, et al. 2011) and in real-time mode (Zhang et al. 2011). This study provides an account of the performance of HWRFX for 87 cases of Atlantic tropical cyclones from the 2005, 2007, and 2009 hurricane seasons.

3. Overview of Atlantic tropical cyclones used in this study

The first step in the HFIP effort was to quantify the impact of increasing the horizontal grid resolution in numerical models on intensity forecasting. In early 2009, HFIP established the High Resolution Hurricane (HRH) tests. The plan for these tests was developed jointly by several segments of the community, including specialists in hurricane forecasting, numerical modeling, and forecast verification. Hurricane specialists at NHC selected 69 cases from 10 hurricanes during the 2005 (Emily, Katrina, Ophelia, Philippe, Rita, and Wilma) and 2007 (Felix, Humberto, Ingrid, and Karen) seasons for the HRH tests. The 69 HRH cases were selected in order to examine a variety of aspects of the models' performance for areas such as 1) the capability to simulate RI events (e.g., Katrina, Rita, Wilma, Humberto, and Karen; see Table 1 and Fig. 2), 2) the impact of vertical wind shear on intensity evolution (e.g., Katrina, Philippe, Rita, Ingrid, and Karen), 3) the effects of eyewall replacement (e.g., Emily, Rita, Wilma, and Felix), 4) the effects of oceanic heating and cooling (e.g., Ophelia), and 5) the impact of terrain interaction during landfall on track and intensity predictions (e.g., Katrina, Rita, and Wilma).

The 2009 hurricane season was a challenging one for forecasts due, in part, to several weak and/or sheared storms (e.g., Ana, Danny, and Erica). As part of the HFIP real-time demonstration system, HWRFX was run in real time during the 2009 season. For this study, cases from five 2009 storms (Ana, Bill, Danny, Erica, and Ida) were added to the 2005 and 2007[1] HRH storm sample to test the HWRFX system's sensitivity to resolution, initial conditions, and physics. In all, 87 cases were tested (Table 1; Fig. 2). Although this is a modest number of cases compared to verification samples used for operational purposes, the sample size should be sufficient for the sensitivity tests presented here. These results will hopefully provide insights that will later be translated into improvements for operational forecasting. The 87 cases composed a diverse sample set of initial intensities, strengthening and weakening scenarios (including a number of RI events), and track types (see Table 1 and Fig. 2) in order to better understand the impact of various factors upon model performance. This was also the first time that the sensitivity of a version of the HWRF model to initial conditions and physics, in addition to resolution, was examined for a comprehensive list of cases.

[1] For this study, cases from Felix (2007) could not be included in the homogeneous sample since all of the necessary initial conditions were not available for that storm.

4. Resolution, configuration, vortex initialization, and model physics

Figure 1 shows the two model configurations (resolutions) used in this study. The domain of simulation along the horizontal direction was set to about $50° \times 50°$ with a moving nest of about $7° \times 7°$. There were 40 hybrid levels in the vertical with the top level set to 50 hPa. The simulations reported here were performed with two kinds of horizontal resolution, namely, (i) a parent domain and moving nest with resolutions of about 27 and 9 km, respectively (27:9 henceforth; Fig. 1a), and (ii) a parent domain and moving nest with resolutions of 9 and 3 km, respectively (9:3 henceforth; Fig. 1b). While the former was selected to be consistent with the current operational resolution, the latter could become the possible operational resolution in the near future. All results reported in this study used moving nests at 9- (i.e., 27:9) or 3-km (i.e., 9:3) resolution. For each case, the model was run to produce 126 h of forecast. As mentioned earlier, the HWRFX system has an option to be initialized using either the operational HWRF gridded binary (GRIB) or GFDL GRIB products. The WRF Preprocessing System (WPS) was used to initialize the model with either the HWRF or the GFDL vortex initializations. The GFS forecasts were used to produce boundary conditions for all the cases reported here.

A description of the operational HWRF suite of physics was recently published as a technical document at the DTC (Gopalakrishnan et al. 2010). The HWRFX physics options used in this study were configured as close as possible to those of the operational HWRF system. The GFDL surface (Bender et al. 2007) and GFS boundary layer formulations (Hong and Pan 1996) were used to parameterize the flux transport and subsequent mixing in the atmosphere. The Ferrier scheme (Ferrier et al. 2002) was used to provide latent heating due to the microphysical processes in the atmosphere, and the simplified Arakawa–Schubert scheme (SAS) was used to parameterize subgrid cumulus-cloud activity. The scheme included a term for momentum mixing (Hong and Pan 1998) that was parameterized as a drag term in the model. Cumulus parameterization in combination with the Ferrier microphysical scheme have been found to have some value in the operational NMM for scales down to about 5 km. However, the contribution to heating from SAS has diminishing returns, that is, an increase in resolution while the grid volume in the inner-core region quickly becomes saturated with the use of an explicit microphysics scheme. Consequently, the SAS convection scheme was switched off at the 3-km resolution (moving mesh) in this work, consistent with Zhang et al. (2011). Finally, to keep the radiation option simple, the

TABLE 1. Summary of the 87 TC cases used in this study. Cases in boldface are for initial intensity (maximum surface wind speed) $\geq$33.4 m s^{-1} (i.e., hurricane strength). RI events are defined as having an increase in maximum surface wind speed in 24 h $\geq$ 15.4 m s^{-1} (i.e., $\geq$30 kt) (Kaplan et al. 2010). See Fig. 2 to view the actual tracks and locations of the model initialization times and RI events.

TC	No. of cases	Dates	RI events
Emily, 2005	8	**0000 UTC 13 Jul, 0000 UTC 14 Jul, 0000 UTC 15 Jul, 0000 UTC 16 Jul, 0000 UTC 17 Jul, 0000 UTC 18 Jul, 0000 UTC 19 Jul, 0000 UTC 20 Jul**	0600 UTC 13 Jul–1200 UTC 15 Jul, 1800 UTC 15 Jul–0000 UTC 17 Jul, 0000 UTC 19 Jul–1200 UTC 20 Jul
Katrina, 2005	6	0000 UTC 24 Aug, 0000 UTC 25 Aug, **0000 UTC 26 Aug, 0000 UTC 27 Aug, 0000 UTC 28 Aug, 0000 UTC 29 Aug**	0600 UTC 26 Aug–0000 UTC 29 Aug
Ophelia, 2005	9	1200 UTC 7 Sep, 1200 UTC 8 Sep, 1200 UTC 9 Sep, **1200 UTC 10 Sep,** 1200 UTC 12 Sep, 1200 UTC 13 Sep, **1200 UTC 14 Sep, 1200 UTC 15 Sep**, 1200 UTC 16 Sep	
Philippe, 2005	6	1200 UTC 17 Sep, 1200 UTC 18 Sep, **1200 UTC 19 Sep**, 1200 UTC 20 Sep, 1200 UTC 21 Sep, 1200 UTC 22 Sep	
Rita, 2005	5	0000 UTC 19 Sep, 0000 UTC 20 Sep, **0000 UTC 21 Sep, 0000 UTC 23 Sep, 0000 UTC 24 Sep**	0000 UTC 20 Sep–0600 UTC 22 Sep
Wilma, 2005	7	0000 UTC 16 Oct, 0000 UTC 17 Oct, 0000 UTC 18 Oct, **0000 UTC 19 Oct, 0000 UTC 22 Oct, 0000 UTC 23 Oct, 0000 UTC 24 Oct**	1800 UTC 17 Oct–1800 UTC 19 Oct
Humberto, 2007	2	1200 UTC 12 Sep, 0000 UTC 13 Sep	0600 UTC 12 Sep–1200 UTC 13 Sep
Ingrid, 2007	4	1200 UTC 12 Sep, 1200 UTC 13 Sep, 1200 UTC 14 Sep, 1200 UTC 15 Sep	
Karen, 2007	3	0000 UTC 25 Sep, 0000 UTC 27 Sep, 0000 UTC 28 Sep	1200 UTC 25 Sep–1800 UTC 26 Sep
Ana, 2009	2	1200 UTC 15 Aug, 0000 UTC 16 Aug	
Bill, 2009	15	0000 UTC 16 Aug, 1200 UTC 16 Aug, 0000 UTC 17 Aug, **1200 UTC 17 Aug, 0000 UTC 18 Aug, 1200 UTC 18 Aug, 0000 UTC 19 Aug, 1200 UTC 19 Aug, 0000 UTC 20 Aug, 1200 UTC 20 Aug, 0000 UTC 21 Aug, 1200 UTC 21 Aug, 0000 UTC 22 Aug, 1200 UTC 22 Aug, 0000 UTC 23 Aug**	
Danny, 2009	5	1200 UTC 26 Aug, 0000 UTC 27 Aug, 1200 UTC 27 Aug, 0000 UTC 28 Aug, 1200 UTC 28 Aug	
Erika, 2009	4	0000 UTC 2 Sep, 1200 UTC 2 Sep, 0000 UTC 3 Sep, 1200 UTC 3 Sep	
Ida, 2009	11	0000 UTC 5 Nov, **1200 UTC 5 Nov,** 0000 UTC 6 Nov, 1200 UTC 6 Nov, 0000 UTC 7 Nov, 1200 UTC 7 Nov, **0000 UTC 8 Nov, 1200 UTC 8 Nov, 0000 UTC 9 Nov,** 1200 UTC 9 Nov, 0000 UTC 10 Nov	0600 UTC 4 Nov–1200 UTC 5 Nov, 1800 UTC 6 Nov–1200 UTC 8 Nov

effect of radiation on the atmosphere was approximated with the NCAR longwave and shortwave radiation scheme (Dudhia's shortwave radiation approach and the Rapid Radiative Transfer Model) available within the WRF framework. Detailed descriptions of the physics options in HWRFX with appropriate references are reported in Yeh et al. (2012). Although the physics options in HWRFX were kept as close as possible to the operational HWRF physics, apart from the difference in the radiation scheme, there were several subtle but significant changes adopted in this study. It may be worthwhile to contrast these differences at least for the 27:9 version. Table 2 specifies some main differences in physics between the 27:9 HWRFX forecasts and the operational HWRF physics.

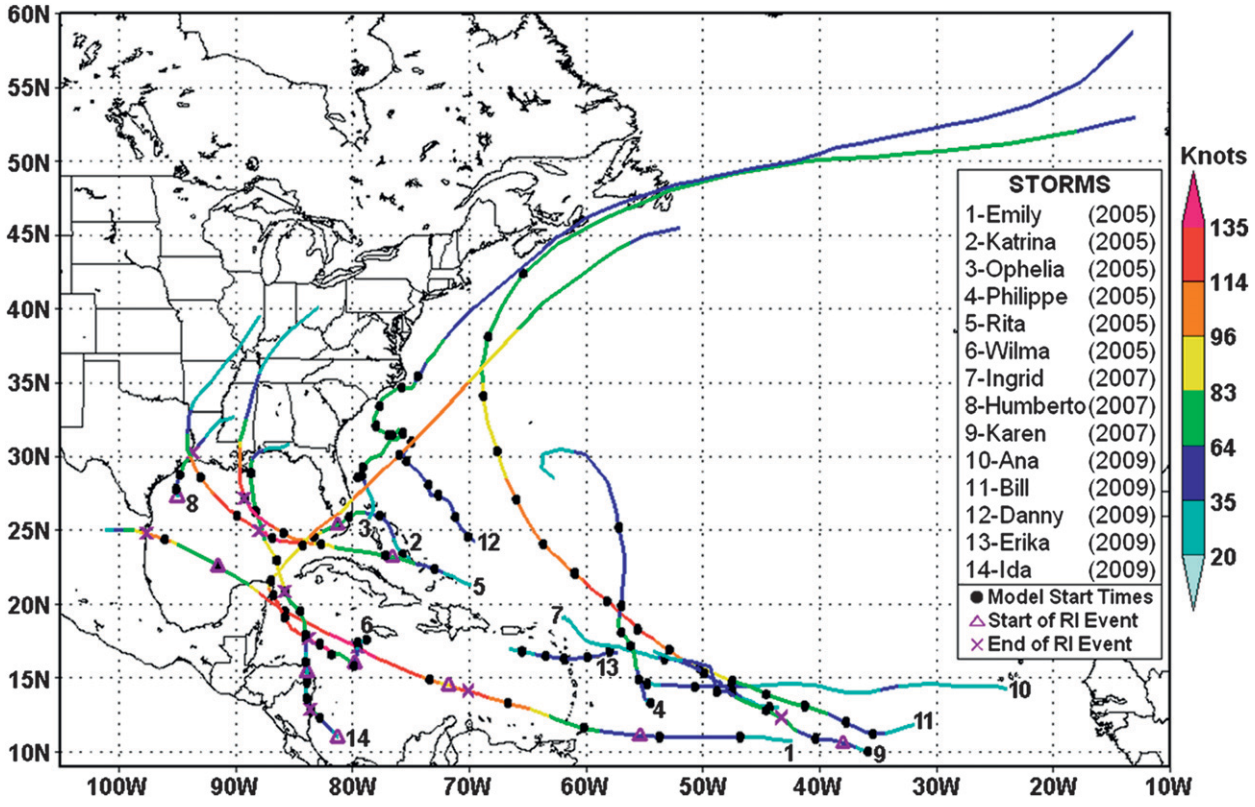

FIG. 2. Actual tracks and intensities (from NHC best-track data) of the 14 tropical cyclones used in this study. Model initialization times (see Table 2) for the 87 cases are indicated by solid dots. Best-track maximum surface (10 m) wind speeds are color coded (speeds on legend are in kt, where 1 kt = 0.514 m s^{-1}). Beginning and ending points of observed RI events are indicated by triangles (△) and exes (×), respectively.

The WRF provides an option for several combinations of physics packages (http://www.mmm.ucar.edu/wrf/users/tutorial/200707/WRF_Physics_Dudhia.pdf). In the recent past, a number of studies have reported intercomparisons between these packages for the specific case of a hurricane (e.g., Li and Pu 2009; Fierro et al. 2009) or for a range of scale-spanning problems such as warm season regional forecasts (e.g., Gallus and Bresch 2006) and hurricane forecasting (Davis et al. 2011). The Kain–Fritsch (KF; Kain and Fritsch 1992) convective scheme, the Yonsei University (YSU; Noh et al. 2003) planetary boundary layer (PBL) scheme, and the WRF-single-moment five-class microphysics scheme (WSM5) form a well-documented combination for hurricane applications (Davis et al. 2011). In this study, the YSU–WSM5–KF combination was used as the second option for testing. The effect of solar radiation on the atmosphere was approximated with the NCAR long- and shortwave radiation scheme available within the WRF framework for all of the runs. Both physics packages in this study used the Noah land surface scheme (Ek et al. 2003). Finally, in all of the runs, sea surface temperatures (SSTs) were obtained from the GFS and held fixed during

TABLE 2. List of differences between operational HWRF and HWRFX physics for the configuration at 27:9.

	Convection	Microphysics	Boundary and surface layer	Land surface and ocean coupling	Dissipative heating	Radiation
Operational HWRF (WRF version 2.0)	SAS with momentum mixing	Ferrier	GFS boundary layer and GFDL hurricane model surface layer scheme (using modifications to surface roughness over ocean)	GFDL slab land surface and POM coupling	Switched on	GFDL radiation scheme
HWRFX (WRF version 3.1; low-resolution forecasts)	SAS without momentum mixing	Ferrier	GFS boundary layer and GFDL hurricane model surface layer scheme (using the original Charnock approximation for surface roughness over ocean)	Noah LSM and uncoupled/ initially prescribed SST	Switched off	NCAR

TABLE 3. List of experiments performed, each containing 87 cases.

No.	Initial conditions	Grid configuration	Physics	Lateral diffusion coefficient in parent and nest domain	Specification
1	HWRF	27:9	HWRFX	1.6 and 0.7	H9hwrf
2	GFDL	27:9	HWRFX	1.6 and 0.7	H9gfdl
3	HWRF	9:3	HWRFX	1.6 and 0.7	H3hwrf
4	GFDL	9:3	HWRFX	1.6 and 0.7	H3gfdl
5	HWRF	9:3	HWRFX	1.6 and 5.0	H3Chwrf
6	HWRF	9:3	NCAR	1.6 and 0.7	H3Nhwrf

the forecasts. Table 3 presents a summary of the sensitivity experiments performed using the 87 cases in this study.

The track and intensity of a storm were determined on the basis of the position of the nest within the parent domain (Fig. 1), which, in turn, is based on the concept of dynamic pressure (Gopalakrishnan et al. 2002). At the end of every time step in the nested domain, the centroid of the dynamic pressure within this moving domain is determined. The minimum dynamic pressure determines the storm center. If the storm center is advected beyond one grid point of the parent domain (three grid points from the center of the nested domain due to the 3:1 parent-to-nest grid ratio), the nested domain is moved to a new position within the parent domain to maintain the storm near the center of the nested domain. The postprocessing software reads the history output file from the moving nest, which carries the information of the minimum sea level pressure, maximum wind, radius of maximum wind, and storm location corresponding to the minimum sea level pressure of the nested domain. As with any numerical model, tracking problems sometimes occurred when the algorithm (or nested grid) lost the storm center, was affected by a storm's proximity to the boundaries, or the nested grid algorithm jumped to a neighboring low pressure system or other storm. These cases were typically removed by screening for situations in which the storm center, as determined from the parent domain and the center determined from the nested grid, differed by more than ~200 km. Once a center position was determined to be suspect, the rest of that forecast (position and intensity) was removed from the sample. Improvements to the automated tracking algorithm are now being developed but were not available when the data for this study were compiled.

5. Results and discussions

Given the large amount of high-resolution data, it is not feasible here to discuss our results in terms of individual cases. Therefore, only statistical verification of the runs will be presented. The results are presented for the homogeneous sample in terms of standard metrics such as (i) absolute track errors and (ii) intensity errors (absolute and bias). These measures provide only a partial glimpse of the high-resolution model's potential to improve track and intensity forecasts. In addition to these standard metrics, this study also examined the radius of maximum wind as a measure of structure prediction. The cumulative distribution function (CDF) of the radius of maximum wind at 10 m above the ground for all the HWRFX forecasts was compared with HRD's H*WIND analyses (e.g., Powell et al. 1998).

Models based solely on climatology and persistence are created from statistical relationships between storm-specific information such as location, time of year, and the behavior of historical storms. For track forecasts, NHC's operational Climatology and Persistence model is CLIPER5 (Neumann 1972). For intensity forecasts, it is DSHIFOR5,[2] which is the Statistical Hurricane Intensity Forecast (SHIFOR5) model (Jarvinen and Neumann 1979; Knaff et al. 2003) adjusted for the inland decay rate of DeMaria et al. (2006) (Decay-SHIFOR). NHC and others (Aberson 1998; Franklin 2010) routinely use these models as a benchmark for establishing the ''skill'' of research and operational models, where CLIPER5 and DSHIFOR5 provide a ''no skill'' baseline. Variations in the errors from these climatology–persistence models help to provide an indication of whether forecast situations are ''easy'' or ''difficult.'' Track forecast results in this study are shown in terms of skill versus CLIPER5. However, because the ultimate focus of this project is on improving intensity forecasts for dynamical models, a higher standard for the intensity forecast baseline, DSHIPS, was used. The Statistical Hurricane Intensity Prediction Scheme (SHIPS) is a sophisticated statistical–dynamical model that predicts

[2] The ''5'' refers to the fact that the original CLIPER (developed in 1972) and SHIFOR (developed in 1979) models only provided forecasts out to 72 h. These forecasts were extended to 120 h (5 days) in 1998 and 2003, respectively.

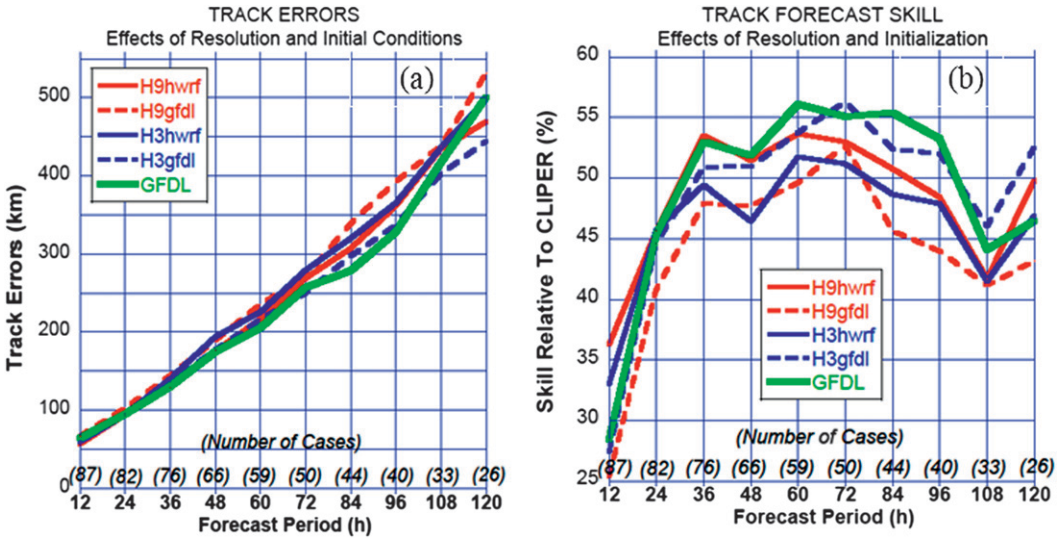

FIG. 3. Verification of HWRFX track forecasts: (a) absolute track errors and (b) skill relative to CLIPER. Here, H9hwrf (H9gfdl) refers to results from the nested domain at 9-km resolution with HWRF (GFDL) initial conditions and H3hwrf (H3gfdl) refers to results from the nested domain at 3-km resolution with HWRF (GFDL) initial conditions.

storm intensity utilizing multiple regression relationships with climatological, persistence, and numerical model predictors (DeMaria et al. 2005). DSHIPS (Decay-SHIPS) is SHIPS adjusted for the decay of storms when they move inland, according to DeMaria et al. (2006), and is regarded by NHC as one of the most reliable intensity forecast models (Franklin 2010). In this study, CLIPER5 and DSHIPS are used as the basis for the comparison of track and intensity predictions, respectively, from HWRFX.[3]

a. Track errors

Figure 3 provides an overview of the track error statistics. The mean track errors for all runs for the two resolutions (i.e., 27:9 and 9:3) and two initial conditions (i.e., GFDL and HWRF) are presented here, where H9hwrf and H9gfdl refer to the 27:9 results with HWRF and GFDL initial conditions, respectively, and H3hwrf and H3gfdl refer to the 9:3 results with HWRF and GFDL initial conditions, respectively (Table 3). The results are also compared with the operational GFDL model. As expected, the number of cases decreases for longer forecast intervals and, consequently, while there are 87 entries for the 12-h forecast, there are only 26 entries for the 120-h forecast. The average track errors for the various models increase almost linearly from near 60 km at 12 h to as low as 436 km (H3gfdl) and as high as 529 km (H9gfdl) at 120 h (Fig. 3a). In general, the large-scale variations that are expected to have a major influence on the TC tracks (Riehl 1954) are well captured by the HWRFX system in all four model versions. The impacts of the changes to the resolution and initial conditions are better seen when the results are normalized with reference to CLIPER5 (Fig. 3b). The operational GFDL model provides the best overall skill. The H3gfdl version of HWRFX has nearly the same skill level overall as the GFDL model and the differences between the average errors from these two models are not statistically significant[4] for any of the forecast periods. Consistent with the results shown in Zhang et al. (2011), use of the

[3] These comparisons are not considered completely "fair" judges of what the performance of the HWRFX (or GFDL) models would be in the actual operational environment since the HWRFX and GFDL versions used here are considered "late" models, i.e., they utilize data from the current operational cycle but finish too late to be available to the hurricane specialists in time to provide guidance for their forecast package. CLIPER5 and DSHIPS are "early" models that finish early enough for the specialists to use their output for the current operational cycle (Franklin 2010). CLIPER5 and DSHIPS both use storm data from the current operational cycle but DSHIPS also utilizes forecasts of the environmental fields from the previous forecast cycle.

[4] For the purpose of this study, statistical significance for the differences between average errors (of two models) is determined by using a Student's *t* test, where the sample size has been adjusted for 24-h serial correlation and an a priori significance threshold of 0.10 (i.e., confidence level of 90%) is used. See Aberson and DeMaria (1994) for a more detailed description of this use of the Student's *t* test and the adjustment for serial correlation. When discussing skill plots, the statistical significance still refers to differences between the average forecast errors.

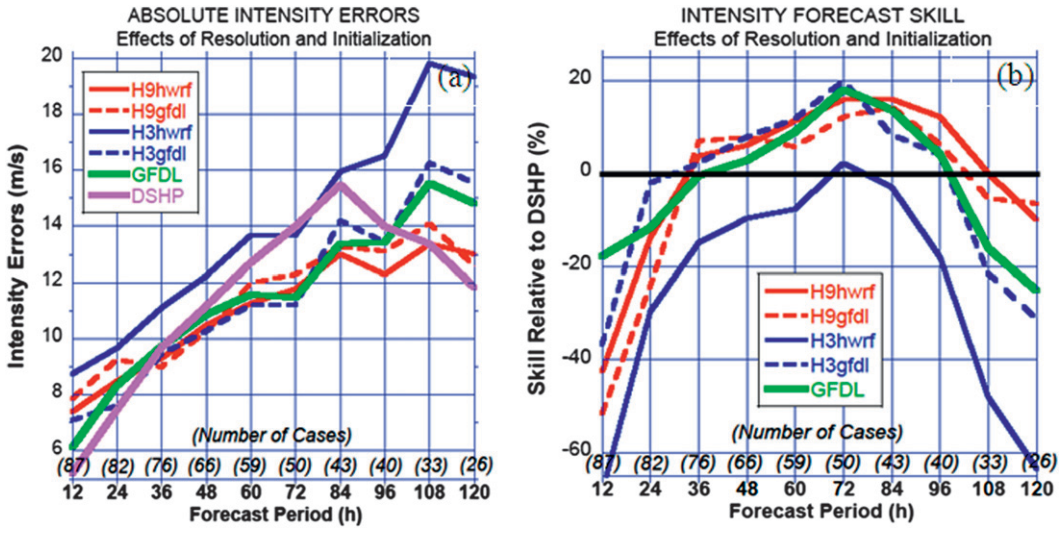

FIG. 4. Verification of HWRFX intensity forecasts: (a) absolute intensity errors and (b) skill relative to DSHIPS. Versions of the HWRFX models are as in Fig. 3.

GFDL initialization in the 9:3 HWRFX version (H3gfdl) produces more skillful results than does the use of the GFDL initialization in the 27:9 version (H9gfdl). These differences in average errors between H3gfdl and H9gfdl are statistically significant at most forecast periods (24, 36, 60, 84, 96, and 120 h). On the contrary, the HWRFX prediction at 27:9 resolution with the corresponding HWRF initialization (H9hwrf) is more skillful in the midforecast intervals compared to the 9:3 version with the HWRF initialization (H3hwrf), though the differences are only statistically significant at 12, 36, and 48 h.

b. *Intensity errors*

Figure 4 depicts the intensity error statistics from HWRFX. The intensity forecasts show a large diversity in behavior (Fig. 4a), with H3hwrf as the outlier with the largest errors. At 12 h, the intensity error for this initialization is close to 9 m s^{-1}, whereas the same initialization with the 27:9 resolution (i.e., H9hwrf) produces an intensity error of about 7.4 m s^{-1}. The differences between H3hwrf and all of the other models are statistically significant at almost all forecast periods, with the exception of DSHIPS from 36 to 96 h. When the average intensity forecast errors are shown as skill relative to DSHIPS (Fig. 4b), the dynamical models are better from at least about 36–96 h, with the exception of H3hwrf, but none of these improvements (versus DSHIPS) are statistically significant except for H3gfdl and GFDL at 72 h. These are encouraging results (see footnote 4) considering DSHIPS is regarded as one of the most reliable intensity forecast models [DSHIPS shows consistent skill relative to DSHIFOR5 (Franklin 2010)]. For this sample, the intensity forecast error results from the various versions of HWRFX (with the exception of H3hwrf) are comparable to or sometimes better than the results from the operational GFDL model for 36–120 h, although none of the differences between the HWRFX models and GFDL are statistically significant.

Considering the impact of the resolution and initial conditions, for this sample there is only a statistically significant improvement at 12 and 24 h using the increased resolution (9:3) with GFDL initial conditions (i.e., H3gfdl versus H9gfdl). However, using the GFDL initial conditions, Zhang et al. (2011) found a more consistent improvement in forecasts with improved resolution. The change in results is likely due to the addition of the sheared and weak cases from 2009, demonstrating the lack of stationarity in verification statistics with these modest sample sizes. On the other hand, runs with the HWRF initial conditions (H9hwrf versus H3hwrf) perform much better at the lower resolution (27:9, i.e., H9hwrf) and these differences are statistically significant at all forecast periods. This is likely caused by the superfluous use of a bogus vortex and an overly restricted storm structure in the operational HWRF initialization. In the operational HWRF system (used here for the initialization of H9hwrf and H3hwrf), weighted information from the first cycle was used in the subsequent cycles (Liu et al. 2006). In fact, weights from the bogus vortex increase with weaker storms. Consequently, the subsequent cycles carry information from the bogus vortex that result in the artificial effects being amplified by the increase in resolution. In summary, the two resolutions provide more or less similar results when using the GFDL initial conditions, at least with the 36–96-h forecasts, but the HWRF initial conditions appear to be much better suited to the coarser grid.

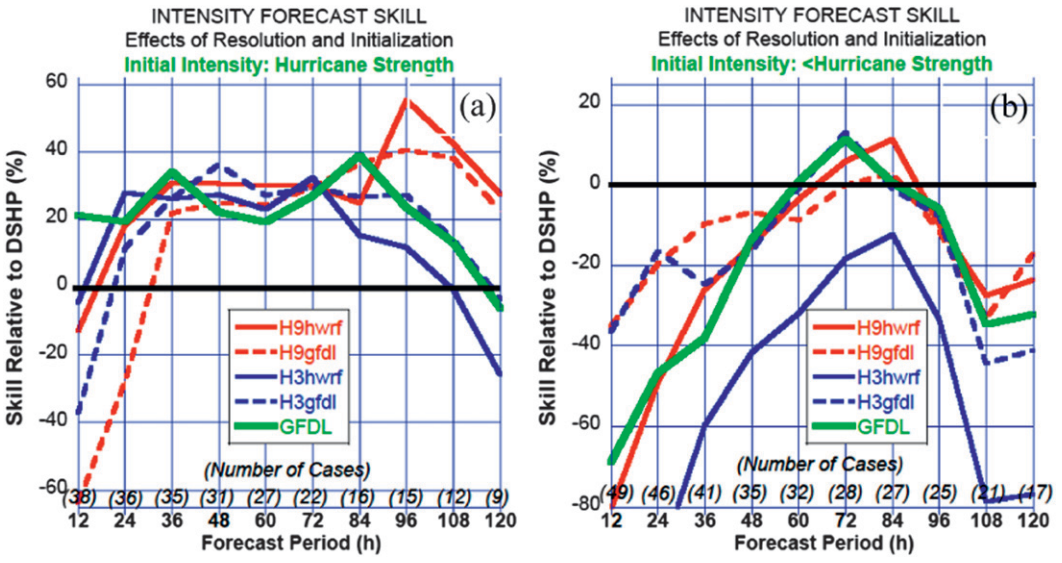

FIG. 5. Skill relative to DSHIPS for (a) storms with initial intensity (maximum sustained surface wind speed) $\geq$33.4 m s^{-1} (i.e., hurricane strength) and (b) storms with initial intensity $<$ 33.4 m s^{-1} (i.e., less than hurricane intensity). See Table 2 for the cases used in (a) and (b).

To further understand the model behavior, the sample was stratified based on the initial intensity of each case. Figures 5a and 5b show the results in terms of skill (versus DSHIPS) for average absolute intensity forecast errors for storms with an initial intensity (maximum sustained surface wind speed) $\geq$ 33.4 m s^{-1} (i.e., hurricane strength) and for storms with an initial intensity $<$33.4 m s^{-1} (i.e., less than hurricane intensity), respectively (see Table 1). Of course, this further reduces the already modest sample size at the 12-h forecast interval to 38 cases for initially stronger storms and 49 cases for initially weaker storms, and the sample size is even smaller for the later forecast intervals. Since the results with DSHIPS for the stratified samples (not shown) demonstrate very little difference for absolute errors between the runs with initially weaker and stronger storms, the baseline used for the skill diagrams is fairly stable. Therefore, the skill differences discussed here for stratification are primarily the result of differences in the model results rather than the baseline.

A comparison between Figs. 5a and 5b demonstrates that the skill in predicting intensity for initially stronger storms far exceeds that of predicting intensity for the initially weaker storms for the given sample. In fact, with the exception of the 12- and 24-h forecast intervals for the HWRFX with GFDL initial conditions (i.e., H9gfdl and H3gfdl), the skill level for every dynamical model represented here is generally higher than 20% from 24 to at least 96 h. With the exception of H3hwrf at 60, 84 and 96 h, virtually all of the positive skill trends (compared to DSHIPS) of the HWRFX models and GFDL are statistically significant from 24 to 96 h, with the skill levels for H9hwrf significant through 108 h and H9gfdl through 120 h. In contrast, the skill for predicting intensity for initially weaker storms was negative for every model at every forecast interval, with the exception of a few models showing slightly positive skill from 60 to 84 h, though none of the positive skill results are statistically significant.

The comparisons used to determine the impacts of increasing the resolution for the stratified samples are mixed. For initially strong storms, the results for runs using the HWRF initialization (H9hwrf and H3hwrf) are fairly close, except for 96 h and beyond when the results with H9hwrf are much better than with H3hwrf and these differences are statistically significant. None of the differences between these two models (for initially strong storms) are significant at early forecast periods. However, for initially weak storms, H9hwrf outperforms H3hwrf at all forecast periods and these differences are all statistically significant. These results demonstrate that the extremely poor performance of the overall sample (Figs. 4a and 4b) of the fine-grid version of HWRFX using HWRF initial conditions (H3hwrf) is mostly due to large errors from the weak storm sample (see Fig. 5a versus Fig. 5b). In fact, H3hwrf performs well out to 72–84 h for the initially strong storm cases but extremely poorly relative to all of the other models depicted for the weaker cases.

In cases for initially stronger storms with GFDL initial conditions (H9gfdl and H3gfdl), the finer resolution has better skill until 72 h, and then the coarser resolution

yields better results from 84 to 120 h. These differences are only statistically significant at 12, 24, and 108 h. However, the sample size was limited to only 22 at 72 h and decreases to only 9 cases at 120 h. For initially weak storms, the results from H9gfdl and H3gfdl are close and any differences are statistically significant only at 36 h.

The comparisons used to determine the impacts of the HWRF versus the GFDL initialization schemes for the stratified samples are also mixed. For initially stronger storms, the HWRF initialization for the 27:9-km HWRFX modeling system (H9hwrf) generally performed better than the GFDL initialization (H9gfdl) at almost all lead times, though these differences are only statistically significant at the 12- and 24-h forecast periods. For initially weaker storms, the only statistically significant differences between H9hwrf and H9gfdl are at the 12- and 24-h forecast periods when the H9gfdl version performed better, though both versions had very negative skill. For the 9:3-km HWRFX modeling system, the GFDL initialization (H3gfdl) performed much better than the HWRF initialization (H3hwrf), with initially weaker storms at all forecast periods and all of these differences are statistically significant except at 84 h. However, for initially stronger storms, H3gfdl and H3hwrf are much closer except at the earlier forecast periods (12–24 h) when H3hwrf performs better and at the later forecast periods (84–120 h) when H3gfdl performs better, but these differences are statistically different only at 12 h. Overall, one would have to conclude that the GFDL initial conditions (H3gfdl) are better suited (for the cases in this study) when using the 9:3-km version, and the HWRF initial conditions (H9hwrf) are better suited when using the 27:9-km version.

Results for the actual absolute intensity errors for initially strong and weak storms and the full sample of cases (not shown) were also very consistent. The smallest errors are associated with initially strong storms at all forecast intervals and show that these stratified results are not biased by scaling with the DSHIPS predictions. Stratification of the track forecast errors (not shown) also yielded very similar results, with skill much higher for initially stronger storms.

c. *Intensity bias*

Figure 6 compares the intensity bias of the models (four versions of HWRFX, GFDL, and DSHIPS). Figures 6a–c show the biases for the complete set of runs (87 cases at 12 h), initially strong storms (38 cases at 12 h), and initially weak storms (49 cases at 12 h), respectively. Several features are worth noting.

(i) For the complete sample (Fig. 6a), the 27:9-km runs with GFDL initialization (H9gfdl) produced

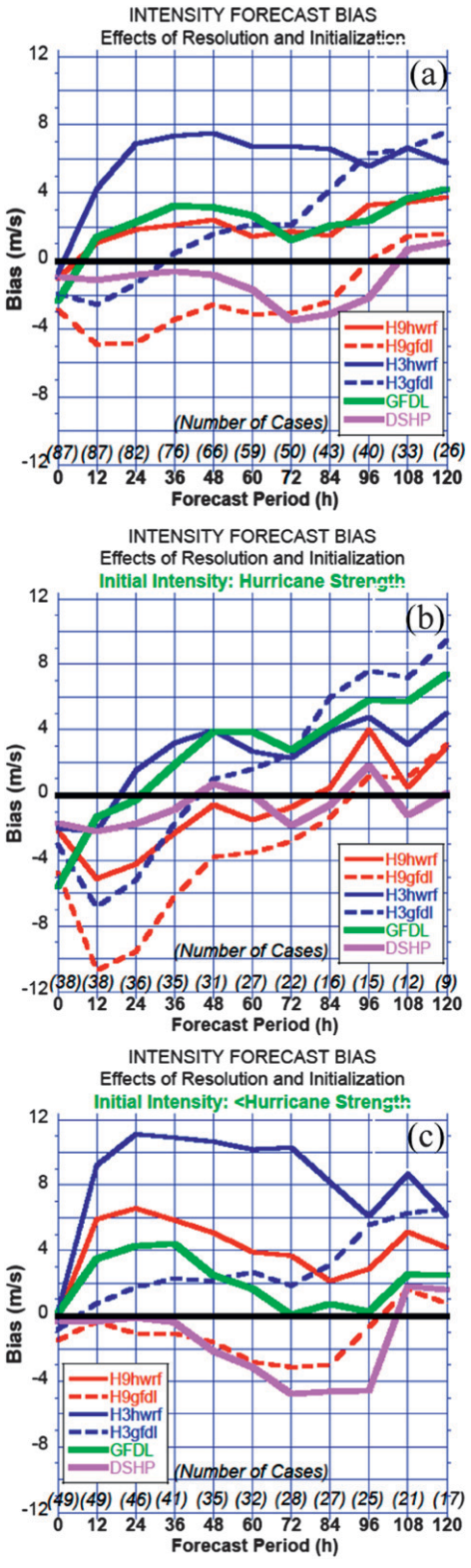

FIG. 6. Intensity bias for the models (four versions of HWRFX, GFDL, and DSHIPS). (a)–(c) Biases for all of the runs (87 cases at 12 h), initially strong storms (38 cases at 12 h), and initially weak storms (49 cases at 12 h), respectively.

noticeable negative bias at lead times through 84 h. This negative bias is primarily from the initially strong storm cases (Fig. 6b) through about 60 h and mainly from the weaker cases (Fig. 6c) at 84 h. The results also show a large negative bias for the 9:3 runs with GFDL initialization (H3gfdl) for the initially stronger storms (Fig. 6b) through 24 h and as with H9gfdl, this negative bias is from the initially strong storm cases. Yeh et al. (2012) used the GFDL initialization for HWRFX at 27:9 and studied the performance of the modeling system for the 2008 season. The study concluded that the early negative bias of the HWRFX resulted from a dynamical inconsistency due to the use of GFDL initial conditions in the HWRFX system, but the continuing negative bias was probably related to inadequate physical forcing of the given configuration (27:9). (Note that the operational GFDL model does not exhibit a strong negative bias for the initially stronger storms except at 0 h.) The study also suggested the need for model-consistent initial conditions for the HWRFX system. The use of initial conditions from the operational HWRF system is hence justified. In the present study, the use of HWRF initial conditions for the HWRFX runs at the same resolution (i.e., 27:9) produced an improved intensity bias and a better intensity forecast (Fig. 5b). ''Downscaling'' (i.e., using the operational HWRF initial conditions from the 27:9 version to drive the experimental 9:3 version: H3hwrf) produced noticeable positive bias (Fig. 6a). The positive bias, however, is almost entirely due to errors from the initially weak storm cases (Fig. 6c).

(ii) The large negative intensity biases of H9gfdl and H3gfdl for the initially strong storm cases during the initial forecast periods (Fig. 6b) develop between the initial time (0 h) and 12 h. This suggests that with the GFDL initial conditions in the 27:9-km HWRFX, a strong initial vortex invariably spins down in the first 12 h of the forecast. This is confirmed by the fact that for H9gfdl the negative bias was very consistent in the initially strong storm sample with the intensity errors being negative at the 12-h forecast period for 87% of the cases and several of these errors being from -25 to -35 m s^{-1}. It takes 24–36 h for the vortex to recover from this spindown, and this is partly reflected in the poor model skill for the same period (Fig. 5a). Intensity bias (Fig. 6b) for each model is strongly associated (or correlated) with model performance for the initially strong storms (Fig. 5a). The GFDL initialization for the 9:3-km runs (H3gfdl) showed a noticeable spindown that lasted for less than 24 h (Fig. 6b). Similar to the results from H9gfdl, the intensity errors are negative at the 12-h forecast period for 74% of the cases. The H3gfdl model consequently recovers from the initial shock to produce more skillful guidance than DSHIPS by 24 h (Fig. 5a). There is a systematic improvement (i.e., reduction) in the initial negative bias with resolution for stronger storms at later lead times (Fig. 6b).

(iii) A careful analysis of the first hour of the simulations (not shown) indicated that the model spindown problem is only marginally affected by the initial intensity errors. For example, the initial (0 h) intensity bias from the cases with initially strong storms (Fig. 6b) for H9gfdl is -4.7 m s^{-1}, for H3gfdl it is -3.0 m s^{-1}, H9hwrf is -2.2 m s^{-1}, H3hwrf is -2.0 m s^{-1}, DSHIPS[5] was 1.7 m s^{-1}, and GFDL is -5.6 m s^{-1}. However, the 12-h (and later) intensity biases for GFDL show that despite a very large 0-h intensity bias the GFDL model recovers rapidly from the initial error and has one of the smallest (negative) biases at 12 h. However, as mentioned earlier, H9gfdl and H3gfdl (and even H9hwrf) spin down in the initially strong storm cases and exhibit relatively large negative intensity biases (-10.7 and -6.8 m s^{-1}, respectively) by the 12-h forecast period. Analyses of the forecasts at earlier (before 12 h) forecast periods (not shown) indicate that much of this spindown occurs in the first hour of the model forecasts. Additional research is needed to examine this vortex initialization problem.

(iv) Despite very small initial (0 h) errors, and small initial biases, cases with initially weak storms presented special challenges for intensity prediction in this study (Fig. 6c). The HWRFX runs with HWRF initialization (H9hwrf and H3hwrf) show substantial positive biases even at 12 h, and the positive biases persisted through 120 h. The worst biases were seen for H3hwrf and are reflected in the very poor intensity forecasts (Fig. 5b). For the cases with initially weak storms at the 12-h forecast period, H3hwrf and H9hwrf forecast errors exhibit a positive bias in 92% and 82% of the cases, respectively, strongly suggesting that these models are cyclogenic in behavior. On the contrary, DSHIPS, which showed negligible bias for the overall and strong

[5] A portion of the initial (0 h) errors for the numerical models can be attributed to the differences between the operational initial intensities and the postprocessed best-track initial intensities. The initial intensities for DSHIPS runs are the same as the actual operational initial intensities. Therefore, the initial DSHIPS errors (and biases) are totally attributed to the errors from the operational initial intensity.

case samples (Figs. 6a and 6b, respectively), showed negligible bias for the weak cases through 36 h but then maintained a negative bias throughout the rest of the forecast period, especially from 60 to 96 h, suggesting that DSHIPS is more conservative than several of the numerical models in intensifying the weaker systems. While improved initial conditions for weak and sheared storms may be of pivotal importance in improving high-resolution model forecasting, it appears that improved physics, especially for longer forecast lead times, may also be an important issue. Some preliminary sensitivity experiments are explored in section 6.

d. *Preliminary evaluation of structure*

The radius of maximum wind (RMW) is defined as the distance between the center of a TC and its band of strongest wind. It is considered an important parameter in TC dynamics and forecasting. In a recent study using idealized initial conditions, Gopalakrishnan et al. (2011) demonstrated that changing the initial radius of maximum wind significantly altered the structure and intensity of the modeled storms. Yet, RMW is one of the parameters with large measurement uncertainties, except where aircraft reconnaissance data are available. As progress is made toward improving structure and intensity predictions with high-resolution models, the need arises to evaluate RMW from the model outputs. Recently, Zhang et al. (2011) used HRD's H*WIND analysis to evaluate the predictions from HWRFX. The same analysis procedure is followed here. Figure 7 provides the cumulative distribution function (CDF) of the RMW[6] from the HWRFX forecasts with the probability counted in 1-km bins of RMW, compared with those from the H*WIND analysis. Excluding samples in which the storm center was located over land, 230 H*WIND analyses, along with 2653 GFDL and 2966 HWRF initialized and 2704 GFDL and 2865 HWRF initialized low- and high-resolution model samples, were used, respectively, to calculate the distribution functions. Figure 7a shows that, according to the H*WIND analysis, 60% (from the 20th to 80th percentiles) of the observed RMWs are distributed over 20–64 km. For the high- and low-resolution HWRFX configurations using GFDL initial conditions (H3gfdl versus H9gfdl; Fig. 7a), 60% of the simulated RMWs are distributed over 34–93 and 43–104 km, respectively. This indicates that storm

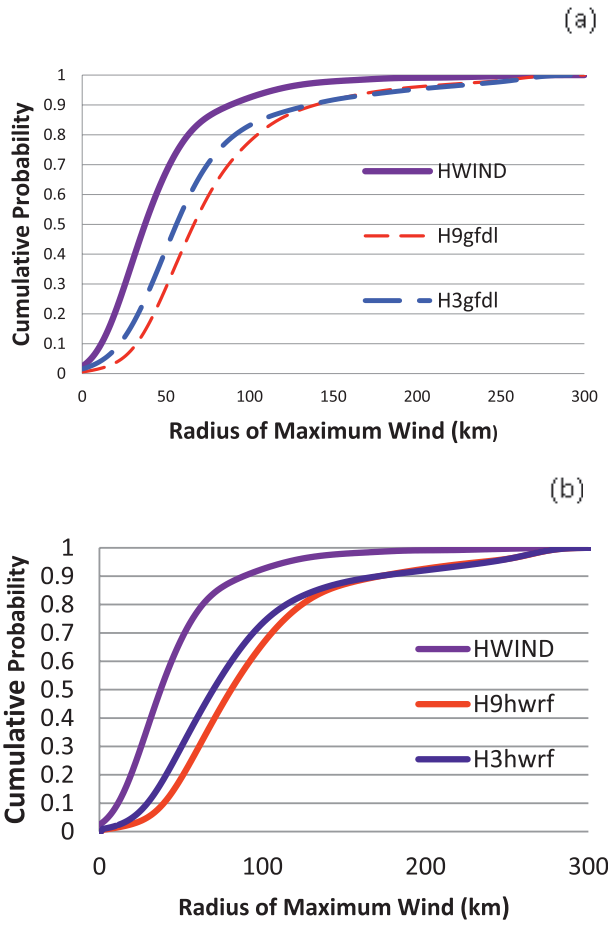

FIG. 7. CDF of the radius of maximum wind at 10 m above the ground for the HWRFX forecasts using (a) H9gfdl and H3gfdl and (b) H9hwrf and H3hwrf, compared with the H*WIND analysis.

size is better predicted with the high-resolution forecasts than with the low-resolution forecasts. Consistent with earlier results, there is a small amount of degradation with the use of HWRF initial conditions (Fig. 7b). However, when compared to the intensity errors produced at higher resolution using the HWRF initial conditions (H3hwrf versus H9hwrf), there is a noticeable improvement in predictions of RMW with increased resolution, even with the HWRF initial conditions (H3hwrf).

6. Additional sensitivity experiments

The operational HWRF system is likely to be run at about 3-km resolution beginning with the 2012 hurricane season using the advanced, model consistent initialization procedure described in section 4. To complement the HWRF system and provide input for further improvements to that system at 3-km resolution, sensitivity experiments were performed on its high-resolution prototype (HWRFX). Specifically, model sensitivity to the model

[6] The model RMWs were obtained from the hourly model output of the azimuthally averaged winds.

physics suite and lateral diffusion coefficient were examined.

a. Sensitivity to model physics

As mentioned in section 4, while the GFS–FERRIER–SAS combination for surface and boundary layer, microphysics, and cumulus convection parameterization scheme is used in the operational HWRF model, an NCAR suite, YSU–WSM5–KF, is another well-researched combination and is used in the Advanced Research core of the WRF (ARW) model. However, it remains a great challenge for the research community to reach a consensus on whether the current physics parameterizations in the operational NWP models are suitable for horizontal grid spacing of 3 km or smaller. Using the physics packages reported in Table 2, Gopalakrishnan et al. (2011) performed a series of idealized experiments with HWRFX at the resolutions down to 3 km, a resolution proposed to be used in NOAA operations. They found that the GFS–FERRIER–SAS combination may be extended to a higher resolution and produced realistic intensification of an isolated hurricane vortex at 3-km resolution as well. Nevertheless, J.-W. Bao et al. (2011, personal communication) more recently performed a series of idealized experiments with HWRFX and found significant sensitivity of the vortex intensification process to the boundary layer parameterization convective scheme followed by the microphysics scheme. They also found that the GFS boundary layer scheme produced deeper inflow and larger storms. These studies are complemented here by examining the sensitivity of the TC tracks and intensity statistics for real cases (Table 1) to the model physics. In this case, the 9:3 configuration of HWRFX was run with HWRF initial conditions but with the physics options set to the YSU–WSM5–KF combination.

Figure 8a provides the track forecast skill obtained from using the modified NCAR (YSU–WSM5–KF) physics combination (H3Nhwrf). There is only a slight difference in the results between the new combination (H3Nhwrf) and the GFS–FERRIER–SAS package (H3hwrf), except at 120 h when there is a 5%–7% improvement in skill, though that difference is not statistically significant. Nevertheless, YSU–WSM5–KF (H3Nhwrf) produces statistically significant improvements in the overall intensity statistics versus the 3-km run with HWRF initial conditions and GFS–FERRIER–SAS physics (H3hwrf) (Fig. 8b) at all forecast intervals except for 12, 24, 84, and 96 h. The H3Nhwrf results are even comparable with the 3-km run with GFDL initial conditions (H3gfdl), the 9-km run with GFDL and with HWRF initial conditions (H9gfdl and H9hwrf, respectively) with GFS–FERRIER–SAS physics (see Fig. 4b), and the operational GFDL model. In fact, for the intensity errors in this sample, the only forecast periods when the GFDL model performs more skillfully than H3Nhwrf (and the differences are statistically significant) are at the 12- and 84-h forecast periods. When the sample is stratified by initially weak and strong storms (Figs. 8c and 8d), the YSU–WSM5–KF combination in HWRFX (H3Nhwrf) produces modest improvements versus H3hwrf for the initially strong vortex cases (Fig. 8c), though these improvements are only statistically significant at 108 h. In the case of initially weak storms, the YSU–WSM5–KF combination (H3Nhwrf) produces large intensity skill improvements at all forecast intervals (on the order of 20%–40%) versus the 3-km HWRF runs with the GFS–FERRIER–SAS combination (H3hwrf), with improvements that are statistically significant between 48 and 72 h. In addition, the H3Nhwrf results are very close to the GFDL forecasts from 36 to 72 h. The intensity error bias for H3Nhwrf (not shown) is just as strongly positive (cyclogenic) at 12 h as H3hwrf, but the bias decreases and becomes negative from 84 to 120 h. A careful reexamination of these results in terms of structure predictions should be performed to better understand the source and nature of the improvements.

b. Sensitivity to lateral diffusion

Apart from the sensitivity to the modeled physical process discussed above, recent work by Rotunno et al. (2009) illustrates the importance of horizontal diffusion, especially in the eyewall region of hurricanes. Nevertheless, the application of lateral diffusion in atmospheric models in general and hurricane models in particular has always been a subject of debate since it is not clear to what extent one can model horizontal diffusion. At this time, diffusion along the horizontal direction is reduced to a minimum value to alleviate the numerical discretization issue (Janjić 1990). In the HWRF/HWRFX system that uses the NMM core as its basis, lateral diffusion is formulated following the Smagorinsky-type nonlinear approach (Janjić 1990), where the horizontal diffusion coefficient is defined as K_h = (Smagorinsky constant $\times$ minimum grid length $\times$ diffusion strength). The Smagorinsky constant is usually set as a tunable parameter. Apart from horizontal deformation, Janjić (1990) included the diffusion strength to be a function of turbulent kinetic energy (TKE). The namelist parameter, known as COAC, is also an input parameter and is defined as 10* (Smagorinsky constant)**2. This coefficient was set to 1.6 in the parent domain and 0.7 in the nested domain in the operational HWRF and in all of the earlier experiments as well (Table 3). However, unlike the NMM system, the HWRF and HWRFX systems are operated with the GFS physics package that does not include an additional equation for the TKE. Consequently, one can expect an effective reduction in lateral diffusion with a different

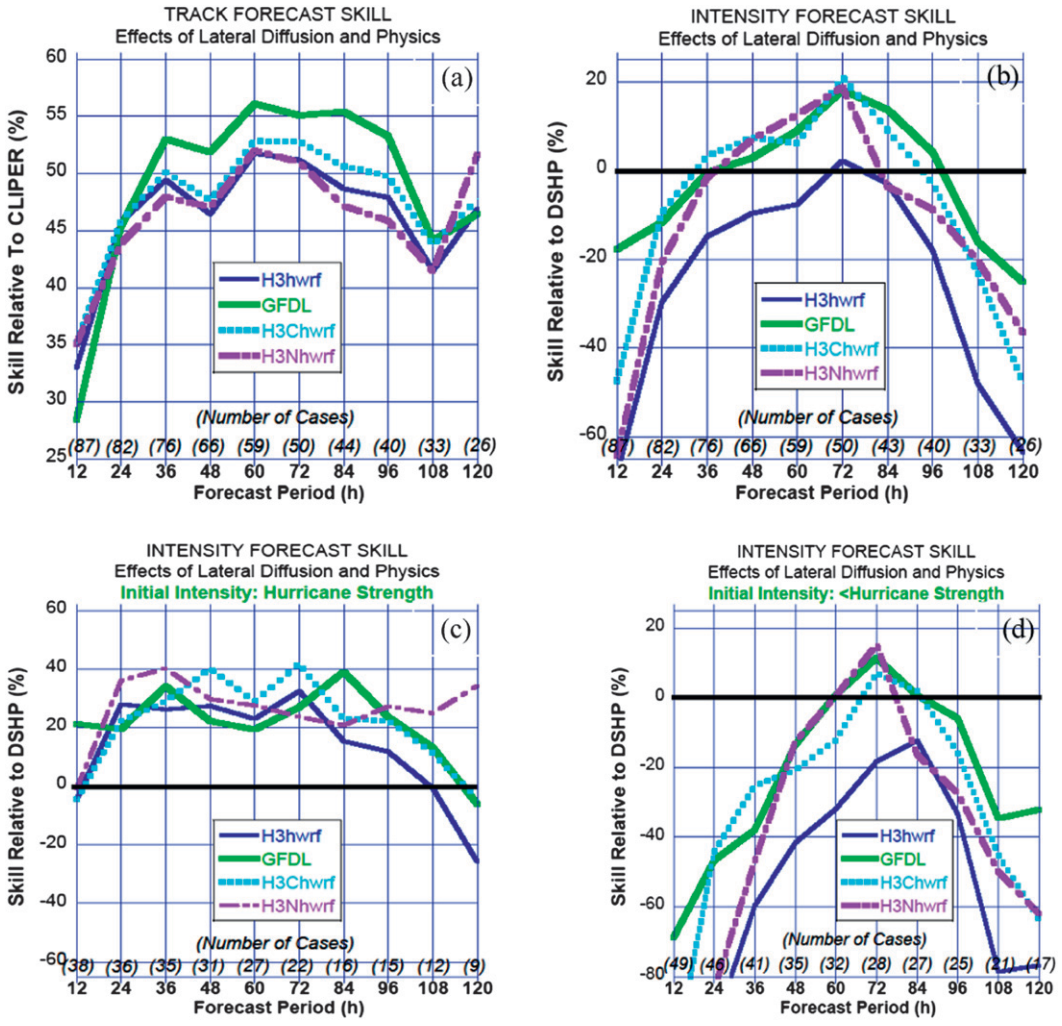

FIG. 8. Verification of HWRFX (a) track errors (skill relative to CLIPER) and intensity errors (skill relative to DSHIPS) for (b) all storms, (c) storms with initial intensity (maximum sustained surface wind speed) ≥ 33.4 m s^{-1} (i.e., hurricane strength), and (d) storms with initial intensity < 33.4 m s^{-1} (i.e., less than hurricane intensity). Effect of the YSU–WSM5–KF physics package (H3Nhwrf) is shown in violet, and enhanced horizontal diffusion on the 3-km HWRFX runs with HWRF initial conditions (H3Chwrf) is shown in cyan.

system of physics that does not include a TKE equation for the NMM core. The COAC was increased to 5 in the inner domain but retained at 1.6 in the outer domain (i.e., increased the lateral diffusion by a factor of roughly 3 in comparison with the parent domain, which is a third of the resolution of the nested domain), and the 9:3 model configuration was rerun, still using the HWRF initial conditions. The standard option of the GFS–FERRIER–SAS physics package was used in this case.

Figure 8 also provides an overview of the sensitivity of the track and intensity statistics to changes in the lateral diffusion coefficient in the 3-km domain. An increase in lateral diffusion in the moving nest (H3Chwrf) produces small improvements to track predictions compared to the H3hwrf results at all forecast intervals (Fig. 8a) and these improvements are statistically significant at 12, 72, 84, and 108 h. Examination of the frequency of superior performance (FSP), a measure (in percent) of how often one model produces a better forecast than another, for the track forecasts from these two models (not shown) also shows these improvements, though small, are fairly consistent. In addition, the H3Chwrf results demonstrate large improvements to the intensity predictions compared with H3hwrf that are statistically significant at all forecast intervals (Fig. 8b). The intensity forecasts with enhanced diffusion (H3Chwrf) are even very close to the GFDL model predictions and, except for at 12 h, there is no statistically significant difference between them. Figures 8c and 8d show the skill results (versus DSHIPS) stratified by initial storm intensity.

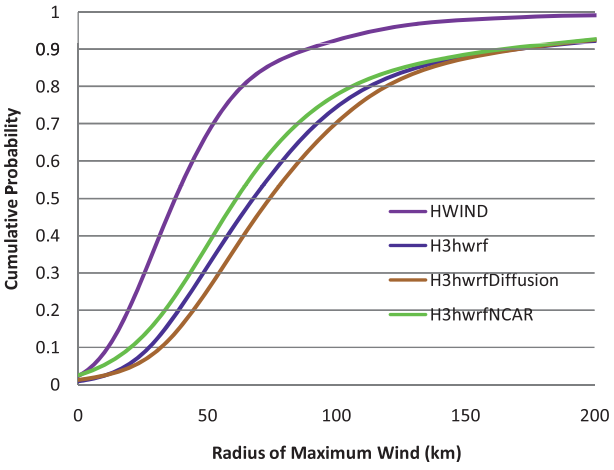

FIG. 9. CDF of the RMW as in Fig. 7, but with H3hwrfDiffusion and H3hwrfNCAR (i.e., H3Chwrf and H3Nhwrf, respectively). For a comparison, H3hwrf is also provided (as in Fig. 7b).

A comparison between H3hwrf and H3Chwrf illustrates that the effect of increasing lateral diffusion produces a small but statistically significant improvement at the 48-, 72-, and 120-h forecast times for the initially strong storms (i.e., hurricane strength). However, improvement for the initially weak storms is large and statistically significant at all forecast times except 120 h, producing results comparable to the GFDL model forecasts at most forecast intervals. The intensity error bias for H3Chwrf (not shown) is still strongly positive (cyclogenic) at 12 h (though not as large as for H3hwrf), but the bias decreases and is minimal from 72 to 120 h. It is very likely that increasing lateral diffusion eases the strong gradient in the eyewall region, especially in the case of initially weak storms and, yet, does not deter the skill on a strong initial vortex (Fig. 8c). This leads to improved skill in the overall intensity predictions with HWRF initial conditions at the 3-km resolution as measured by the current skill metrics.

Figure 9 provides the CDF of the RMW from the HWRFX forecasts with the probability counted in 1-km bins of RMW, compared with those from the H*WIND analysis. Excluding samples in which the storm center was located over land, 230 H*WIND analyses, along with 2966 data points obtained from runs with the YSU–WSM5–KF combination (H3Nhwrf) and runs with changes in the lateral diffusion coefficient in the 3-km domain (H3Chwrf), were used. The baseline from HWRFX (H3hwrf) is also provided for comparison (as in Fig. 7b). Although the increase in lateral diffusion improves the intensity statistics (Fig. 8), the structure, which in this study is measured in terms of the radius of maximum wind, is degraded with the GFS–FERRIER–SAS combination.

7. Summary and conclusions

In this study, the research version of the operational HWRF system, HWRFX, was run for 5 days with 87 TC cases (Table 1) at two horizontal resolutions, namely, (i) a parent domain with a resolution of about 27 km with a 9-km moving nest (i.e., 27:9) and (ii) a parent domain at a resolution of 9 km with a moving nest at 3 km (i.e., 9:3). Two initial conditions were used for each horizontal resolution, namely, the GFDL and operational HWRF initializations. As part of additional sensitivity experiments, an NCAR suite of physics (YSU–WSM5–KF combination) was used with the operational HWRF initial conditions for the fine grid model version (9:3 km). Sensitivity experiments were also performed using the 9:3-km version with HWRF initial conditions by adjusting the diffusion coefficient. This study has shown the following:

1) The 9:3-km runs with the HWRFX system using a system of physics close to the operational version of the HWRF model and the GFDL initial conditions (H3gfdl) provided the best overall skill in terms of both track and intensity predictions. This configuration also produced improved structure predictions, as measured in terms of probability estimates of the radius of maximum wind.
2) The 27:9-km runs with the HWRFX system using a system of physics close to the operational version of the HWRF model and the HWRF initial conditions (H9hwrf) provided reasonable skill in terms of both track and intensity predictions. However, the 9:3-km version with the same configuration (H3hwrf) produced very poor results, showing that the HWRF initial conditions may need adjustment to be consistent with the finer-scale version of the model. Work is currently proceeding in this area.
3) In cases of initially strong storms (hurricane strength), initialization of the hurricane vortex in HWRFX appeared to have a major influence on the intensity bias and errors, at least during the first 36 h of the model run. In general, the model appeared to rapidly spin down the initial vortex for strong storms in a majority of the cases when using the GFDL initial conditions (i.e., H9gfdl and H3gfdl). Use of HWRF initial conditions for the initially stronger storms reduced model bias and improved skill between 0 and 24 h. Nevertheless, the operational GFDL model performed better than any of the other models discussed here for the 12-h intensity forecasts for initially strong storms.
4) Forecasting intensity for initially weak storms poses major problems and requires further study. Other than marginal skill at 72–84 h, none of the models,

including GFDL, had any skill relative to DSHIPS for these cases. The 9:3-km runs using HWRF initialization and to a lesser extent with 27:9 km as well (H3hwrf and H9hwrf, respectively) illustrated a substantial positive bias, indicating the overly cyclogenic behavior of HWRFX for initially weaker storms. This is possibly caused by the superfluous use of a bogus vortex and an overly restricted storm structure in the operational HWRF initialization. In the operational HWRF system (used here for the initialization of H9hwrf and H3hwrf), weighted information from the first cycle was used in the subsequent cycles (Liu et al. 2006). In fact, weights from the bogus increase with weaker storms. Consequently, the subsequent cycles carry information from the bogus vortex that results in the artificial effects being amplified by the increase in resolution. Increasing the lateral diffusion on the 3-km domain for runs initialized with the HWRF initial conditions (H3Chwrf) substantially improved the intensity skill for the initially weak storm cases, leading to skill that equaled that of the GFDL forecasts. These improvements were still marginal, however, when compared to the DSHIPS predictions.

5) Increasing lateral diffusion on the 3-km domain runs with the operational GFS–FERRIER–SAS combination for the surface and boundary layers, microphysics, and the cumulus convection parameterization scheme and HWRF initial conditions (H3Chwrf) had a slight improvement in intensity skill at most forecast times for the initially strong storms (i.e., hurricane strength). Nevertheless, it had nearly the same effect on track and intensity skill as using the NCAR YSU–WSM5–KF physics combination, especially for initially weak storms. However, with the increased lateral diffusion, the structure, which in this study is measured in terms of the radius of maximum wind, is slightly degraded with the GFS–FERRIER–SAS combination. A further analysis beyond the standard metrics adopted here will be the subject of a future work.

The use of high-resolution numerical models for predicting rapid intensification (RI) episodes is in its stages of infancy. Unfortunately, because of the greatly reduced sample size in this study (see Tables 1 and 4), a statistical analysis of RI events may not be feasible. In addition, track error and landfall timings may further dilute any statistical analysis of such events. Moreover, it is likely that one may have to look beyond the 10-m wind to gain important insights into the modeled RI problem. In recent publications (and perhaps one of the few references to RI using advanced models), Rogers (2010) and Chen et al. (2011) provide some extensive analyses

TABLE 4. Observed intensity changes and predicted intensity changes as measured by 10-m wind speed ($m\ s^{-1}$) produced by each of the HWRFX model configurations during each of the observed RI events shown in Table 1. These changes were measured when data were available during the closest forecast period to the RI cycle. For example, for RI between 1800 UTC 15 Jul and 0000 UTC 17 Jul in Emily, the 0000 UTC 14 Jul forecasts were used to obtain the predicted intensity change for the 1800 UTC 15 Jul–0000 UTC 17 Jul time period. In a case like the Humberto RI event (0600 12 Sep–1200 UTC 13 Sep) the closest forecast cycle was 1200 UTC 12 Sep, which only allowed the intensity change to be obtained for 1200 12 Sep–1200 UTC 13 Sep.

TC	RI event	Observed intensity change	Forecast cycle	H9hwrf	H9gfdl	H3hwrf	H3gfdl	H3Chwrf	H3Nhwrf
Emily, 2005	0600 UTC 13 Jul–1200 UTC 15 Jul	36.0	0000 UTC 13 Jul	17.9	22.4	13.3	23.0	−0.5	13.3
	1800 UTC 15 Jul–0000 UTC 17 Jul	23.1	0000 UTC 15 Jul	−7.1	−7.7	−9.2	0	−5.1	−16.8
	0000 UTC 19 Jul–1200 UTC 20 Jul	18.0	0000 UTC 19 Jul	4.1	11.2	−0.5	20.9	3.1	12.2
Katrina, 2005	0600 UTC 26 Aug–0000 UTC 29 Aug	38.6	0000 UTC 26 Aug	25.5	33.7	34.7	38.8	33.2	28.6
Rita, 2005	0000 UTC 20 Sep–0600 UTC 22 Sep	48.9	0000 UTC 20 Sep	22.4	9.2	27.5	25.5	21.9	26.0
Wilma, 2005	1800 UTC 17 Oct–1800 UTC 19 Oct	48.9	0000 UTC 17 Oct	21.4	32.6	24.0	37.7	19.4	24.5
Humberto, 2007	0600 UTC 12 Sep–1200 UTC 13 Sep	20.6	1200 UTC 12 Sep	7.7	1.0	12.2	3.1	7.7	13.3
Karen, 2007	1200 UTC 25 Sep–1800 UTC 26 Sep	15.4	0000 UTC 25 Sep	2.6	10.2	14.8	12.8	7.7	12.2
Ida, 2009	0600 UTC 4 Nov–1200 UTC 5 Nov	20.6	0000 UTC 5 Nov	6.1	4.1	11.2	9.2	9.2	9.2
	1800 UTC 6 Nov–1200 UTC 8 Nov	28.3	0000 UTC 6 Nov	14.3	8.2	35.7	21.9	24.5	23.0

of RI, respectively, of Hurricanes Dennis and Wilma during 2005. An extensive work of the above sort may be required using the current datasets. Table 4 provides the intensity changes as measured by 10-m wind speed produced by each of the model configurations during the observed RI events for the cases in this study. The information provided in Table 4 may be useful for further research.

The research version of the system (HWRFX) is not identical to the operational version (HWRF) due to minor differences in the physics, domain size, methods in which the initial conditions are generated, and ocean coupling. The HWRFV3.2 is an ocean-coupled, high-resolution operational and research system currently under development. This system is a merger between HWRFX and the operational HWRF system. The current work provides a basis for further evaluation of this system. However, as the research and operational communities work together under the auspices of NOAA's HFIP to better understand the degree to which tropical cyclone intensity forecasts can be improved by increasing the horizontal grid spacing of operational numerical weather prediction (NWP) models, it remains a greater challenge for these communities to reach a consensus on the vortex initialization scheme and whether the current physics parameterizations in operational NWP models are even suitable for weak and sheared storms. Apart from the initialization problems for weak systems, the current study also illustrates the importance of diffusion and related boundary layer processes especially in the forecast of weaker storms. It should be noted that the HWRF system of parameterization schemes was chosen to be consistent with the other operational models such as the GFDL and the GFS system and, perhaps, part of the insensitivity on tracks was due to physics best ''tuned'' for those hydrostatic models. As rightly mentioned by one of the reviewers of this manuscript, it is well known that the HWRF system of physics is diffusive. Therefore, increasing the horizontal diffusion, as carried out in the current study, may only clearly provide improvements to intensity skill as measured by the standard metrics used at this time. Also, those improvements may occur at the expense of structure predictions. Nevertheless, thanks largely due to the observations now available in the inner core (Zhang et al. 2011a,b; Zhang and Montgomery 2012), we are in the process of evaluating both the vertical and horizontal diffusivities in the HWRF system. In fact we find that the overly diffusive inner core is due to the vertical diffusion in the HWRF system. These findings have been tested within the idealized paradigm and need to be evaluated for the cases under consideration. As more observations become available under HFIP, we are hoping to evaluate the horizontal diffusivity in both an idealized as well as a real framework. Our subsequent studies will focus on the impact of physics on weak storms at higher resolutions.

Acknowledgments. The authors acknowledge funding from NOAA's Hurricane Forecast Improvement Project that supported this work. The model was primarily developed at NCEP. The first author wishes to acknowledge several of the scientists at NCEP for their help during the development of this modeling system. The authors wish to thank Gail Derr, Lisa Bucci, Drs. Robert Rogers, and Tomislava Vukicevic, as well as the two anonymous reviewers (who provided very detailed reviews), for providing useful comments on improving the original version of this manuscript.

REFERENCES

Aberson, S. D., 1998: Five-day tropical cyclone track forecasts in the North Atlantic basin. *Wea. Forecasting,* **13,** 1005–1015.

——, and M. DeMaria, 1994: Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Wea. Rev.,* **122,** 2804–2815.

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

Chen, H., D.-L. Zhang, J. Carton and R. Atlas, 2011: On the rapid intensification of Hurricane Wilma (2005). Part I: Model prediction and structural changes. *Wea. Forecasting,* **26,** 885–901.

Davis, C., and Coauthors, 2008: Prediction of landfalling hurricanes with the Advanced Hurricane WRF model. *Mon. Wea. Rev.,* **136,** 1990–2005.

——, and Coauthors, 2011: High-resolution hurricane forecasts. *Comput. Sci. Eng.,* **13,** 22–30.

DeMaria, M., and J. Kaplan, 1999: An updated Statistical Hurricane Intensity Prediction Scheme for the Atlantic and eastern North Pacific basins. *Wea. Forecasting,* **14,** 326–337.

——, ——, J. A. Knaff, M. Mainelli, and L. K. Shay, 2005: Further improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS). *Wea. Forecasting,* **20,** 531–543.

——, J. A. Knaff, and J. Kaplan, 2006: On the decay of tropical cyclone winds crossing narrow landmasses. *J. Appl. Meteor. Climatol.,* **45,** 491–499.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the NCEP operational mesoscale Eta Model. *J. Geophys. Res.,* **108,** 8851, doi:10.1029/2002JD003296.

Ferrier, B. S., Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in the NCEP Eta Model. Preprints, *15th Conf. on Numerical Weather Prediction,* San Antonio, TX, Amer. Meteor. Soc., 280–283.

Fierro, A. O., R. F. Rogers, F. D. Marks, and D. S. Nolan, 2009: The impact of horizontal grid spacing on the microphysical and kinematic structures of strong tropical cyclones simulated with the WRF-ARW model. *Mon. Wea. Rev.,* **137,** 3717–3743.

Franklin, J. L., cited 2010: 2009 National Hurricane Center verification report. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2009.pdf.]

Gallus, W. A., Jr., and J. F. Bresch, 2006: Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts. *Mon. Wea. Rev.,* **134,** 2632–2641.

Ginis, I., W. Shen, and M. A. Bender, 1999: Performance evaluation of the GFDL coupled hurricane ocean prediction system in the Atlantic basin. Preprints, *23rd Conf. on Hurricanes and Tropical Meteorology,* Dallas, TX, Amer. Meteor. Soc., 607–610.

Gopalakrishnan, S. G., and Coauthors, 2002: An operational multiscale atmospheric model with grid adaptivity for hurricane forecasting. *Mon. Wea. Rev.,* **130,** 1830–1847.

——, N. Surgi, R. Tuleya, and Z. Janjić, 2006: NCEP's two-way-interactive-moving-nest NMM-WRF modeling system for hurricane forecasting. Preprints, *27th Conf. on Hurricanes and Tropical Meteorology,* Monterey, CA, Amer. Meteor. Soc., 7A.3. [Available online at http://ams.confex.com/ams/pdfpapers/107899.pdf.]

——, Q. Liu, T. Marchok, D. Sheinin, N. Surgi, R. Tuleya, R. Yablonsky, and X. Zhang, 2010: Hurricane Weather and Research and Forecasting (HWRF) model scientific documentation. NOAA/Earth System Research Laboratory, 75 pp.

——, F. D. Marks, X. Zhang, J.-W. Bao, K.-S. Yeh, and R. Atlas, 2011: The experimental HWRF system: A study on the influence of horizontal resolution on the structure and intensity changes in tropical cyclones using an idealized framework. *Mon. Wea. Rev.,* **139,** 1762–1784.

Gross, J. M., 1999: 1998 National Hurricane Center forecast verification. Minutes, *53rd Interdepartmental Hurricane Conf.,* Biloxi, MS, Office of the Federal Coordinator for Meteorological Services and Supporting Research, B24–B63. [Available from Office of the Federal Coordinator for Meteorological Services and Supporting Research, Ste. 1500, 8455 Colesville Rd., Silver Spring, MD 20910.]

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.,* **124,** 2322–2339.

——, and ——, 1998: Convective trigger function for a mass-flux cumulus parameterization scheme. *Mon. Wea. Rev.,* **129,** 1164–1178.

Janjić, Z. I., 1990: The step-mountain coordinates: Physical package. *Mon. Wea. Rev.,* **118,** 1429–1443.

——, 2003: A non-hydrostatic model based on a new approach. *Meteor. Atmos. Phys.,* **82,** 271–285.

——, J. P. Gerrity Jr., and S. Nickovic, 2001: An alternative approach to nonhydrostatic modeling. *Mon. Wea. Rev.,* **129,** 1164–1178.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity for the North Atlantic basin. NOAA Tech. Memo. NWS NHC-10, 22 pp.

Kain, J. S., and J. M. Fritsch, 1992: The role of convective ''trigger function'' in numerical forecasts of mesoscale convective systems. *Meteor. Atmos. Phys.,* **49,** 93–106.

Kaplan, J., M. DeMaria, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting,* **25,** 220–241.

Knaff, J. A., M. DeMaria, B. Sampson, and J. M. Gross, 2003: Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting,* **18,** 80–92.

Kurihara, Y., 1975: Budget analysis of a tropical cyclone simulated in an axisymmetric numerical model. *J. Atmos. Sci.,* **32,** 25–59.

——, and R. E. Tuleya, 1974: Structure of a tropical cyclone developed in a three-dimensional numerical simulation model. *J. Atmos. Sci.,* **31,** 893–919.

——, M. A. Bender, R. E. Tuleya, and R. J. Ross, 1995: Improvements in the GFDL hurricane prediction system. *Mon. Wea. Rev.,* **123,** 2791–2801.

Li, X., and Z. Pu, 2009: Sensitivity of numerical simulation of early rapid intensification of Hurricane Emily (2005) to cumulus parameterization schemes in different model horizontal resolutions. *J. Meteor. Soc. Japan,* **87,** 403–421.

Liu, Q., N. Surgi, S. Lord, W.-S. Wu, S. Parrish, S. Gopalakrishnan, J. Waldrop, and J. Gamache, 2006: Hurricane initialization in HWRF model. Preprints, *27th Conf. on Hurricanes and Tropical Meteorology,* Monterey, CA, Amer. Meteor. Soc., 8A.2. [Available online at http://ams.confex.com/ams/pdfpapers/108496.pdf.]

Liu, Y., D.-L. Zhang, and M. K. Yau, 1997: A multiscale numerical study of Hurricane Andrew (1992). Part I: Explicit simulation and verification. *Mon. Wea. Rev.,* **125,** 3073–3093.

——, ——, and ——, 1999: A multiscale numerical study of Hurricane Andrew (1992). Part II: Kinematics and inner-core structures. *Mon. Wea. Rev.,* **127,** 2597–2616.

Marks, F. D., and R. A. Houze, 1987: Inner core structure of Hurricane Alicia from airborne Doppler radar observations. *J. Atmos. Sci.,* **44,** 1296–1317.

——, R. A. Houze Jr., and J. F. Gamache, 1992: Dual-aircraft investigation of the inner core of Hurricane Norbert. Part I: Kinematic structure. *J. Atmos. Sci.,* **49,** 919–942.

Neumann, C. B., 1972: An alternate to the HURRAN (hurricane analog) tropical cyclone forecast system. NOAA Tech. Memo. NWS SR-62, 24 pp.

Noh, Y., W. G. Cheon, S.-Y. Hong, and S. Raasch, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.,* **107,** 401–427.

Powell, M. D., S. H. Houston, L. R. Amat, and N. Morisseau-Leroy, 1998: The HRD real-time hurricane wind analysis system. *J. Wind Eng. Ind. Aerodyn.,* **77-78,** 53–64.

Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting,* **24,** 395–419.

Riehl, H., 1954: *Tropical Meteorology*. McGraw-Hill, 392 pp.

Rogers, R., 2010: Convective-scale structure and evolution during a cloud-resolving simulation of tropical cyclone rapid intensification. *J. Atmos. Sci.,* **67,** 44–70.

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.,* **90,** 1783–1788.

Tuleya, R. E., and Y. Kurihara, 1975: The energy and angular momentum budgets of a three-dimensional tropical cyclone model. *J. Atmos. Sci.,* **32,** 287–301.

——, and ——, 1978: A numerical simulation of the landfall of tropical cyclones. *J. Atmos. Sci.,* **35,** 242–257.

——, and ——, 1982: A note on the sea surface temperature sensitivity of a numerical model of tropical storm genesis. *Mon. Wea. Rev.,* **110,** 2063–2069.

Willoughby, H. E., 1979: Forced secondary circulations in hurricanes. *J. Geophys. Res.,* **84,** 3173–3183.

——, 1990a: Gradient balance in tropical cyclones. *J. Atmos. Sci.,* **47,** 265–274.

——, 1990b: Temporal changes of the primary circulation in tropical cyclones. *J. Atmos. Sci.,* **47,** 242–264.

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, R. Rogers, F. D. Marks, and R. Atlas, 2012: Performance of the experimental HWRF in the 2008 hurricane season. *Nat. Hazards,* doi:10.1007/s11069-011-9787-7, in press.

Zhang, D.-L., Y. Liu, and M. K. Yau, 1999: Surface winds at landfall of Hurricane Andrew (1992)—A reply. *Mon. Wea. Rev.,* **127,** 1711–1721.

——, ——, and ——, 2000: A multiscale numerical study of Hurricane Andrew (1992). Part III: Dynamically induced vertical motion. *Mon. Wea. Rev.,* **128,** 3772–3788.

——, ——, and ——, 2001: A multiscale numerical study of Hurricane Andrew (1992). Part IV: Unbalanced flows. *Mon. Wea. Rev.,* **129,** 92–107.

——, ——, and ——, 2002: A multiscale numerical study of Hurricane Andrew (1992). Part V: Inner-core thermodynamics. *Mon. Wea. Rev.,* **130,** 2745–2763.

Zhang, J. A., and M. T. Montgomery, 2012: Observational estimates of the horizontal eddy diffusivity and mixing length in the low-level region of intense hurricanes. *J. Atmos. Sci.,* **69,** 1306–1316.

——, F. D. Marks Jr., M. T. Montgomery, and S. Lorsolo, 2011a: An estimation of turbulent characteristics in the low-level region of intense Hurricanes Allen (1980) and Hugo (1989). *Mon. Wea. Rev.,* **139,** 1447–1462.

——, R. F. Rogers, D. S. Nolan, and F. D. Marks, 2011b: On the characteristic height scales of the hurricane boundary layer. *Mon. Wea. Rev.,* **139,** 2523–2535.

Zhang, X., T. S. Quirino, S. Gopalakrishnan, K.-S. Yeh, F. D. Marks Jr., and S. B. Goldenberg, 2011: HWRFX: Improving hurricane forecasts with high resolution modeling. *Comput. Sci. Eng.,* **13,** 13–21.