

Volume 94 Number 3 March 2013

BAMS



Bulletin of the American Meteorological Society

STATEWIDE CLIMATE EXTREMES

HURRICANE INTENSIFICATION STUDY

HIGH-LATITUDE SURFACE FLUXES



**CAN WE FORECAST THIS
ONE WEEK IN ADVANCE?**

THE HURRICANE FORECAST IMPROVEMENT PROJECT

BY ROBERT GALL, JAMES FRANKLIN, FRANK MARKS, EDWARD N. RAPPAPORT, AND FREDERICK TOEPFER

Since 2008, HFIP has engaged the tropical community in hopes of longer lead times and greater accuracy in warnings, resulting in technology-driven savings across the economy.

Tropical cyclone activity in the Atlantic hurricane basin broke records for numbers and impacts during the first decade of the new millennium. A total of 13 hurricanes crossed the contiguous U.S. coastline from 2000 to 2010, including such now infamous storms as Charley (2004), Katrina (2005), Rita (2005), Wilma (2005), and Ike (2008). In 2005 alone, 27 Atlantic systems reached tropical storm status, far surpassing the previous record of 21. The heightened activity brought an increased awareness of the dangers from tropical cyclones and led to a number of studies concerning the National Oceanic and Atmospheric Administration's (NOAA's) ability to forecast hurricanes. The additional attention on the nation's hurricane warning program provided

opportunities to give visibility to and initiate actions on intensity forecasting, a critical area where no appreciable improvement has been made over the preceding two decades (e.g., Cangialosi and Franklin 2011). To address this issue, NOAA, through its Science Advisory Board (SAB), established a Hurricane Intensity Research Working Group (HIRWG), which documented its recommendations to improve forecasts of hurricane intensity in October 2006 (NOAA SAB 2006). In addition, the National Science Foundation (NSF) National Science Board issued a report in January 2007 on the need for a National Hurricane Research Initiative (NSF 2007) and the Office of the Federal Coordinator of Meteorological Services (OFCM) issued a report in February 2007 calling for a federal investment of \$70–\$85 million (all amounts are in U.S. dollars; hereafter dollar amounts in millions are shown in the form \$85M) annually over the next 10 years for tropical cyclone research and development, transition of research to operations, and operational high-performance computing (OFCM 2007).

NOAA's response was the establishment of the Hurricane Forecast Improvement Project (HFIP), as noted in this November 2007 statement: "In response to the HIRWG report, NOAA convened a corporate hurricane summit developing unified strategy to address hurricane forecast improvements. On 10 May 2009 the NOAA Executive Council (NEC) established the NOAA Hurricane Forecast Improvement Project

AFFILIATIONS: GALL AND TOEPFER—NOAA/NWS, Silver Spring, Maryland; FRANKLIN AND RAPPAPORT—NOAA/NWS/National Hurricane Center, Miami, Florida; MARKS—NOAA/OAR/AOML/Hurricane Research Division, Miami, Florida

CORRESPONDING AUTHOR: Robert Gall, National Oceanic and Atmospheric Administration, 1325 East-West Highway, Silver Spring, MD 20910
E-mail: robert.gall@noaa.gov

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-12-00071.1

In final form 8 August 2012
©2013 American Meteorological Society

(HFIP), a 10-year effort to accelerate improvements in 1–5-day forecasts for hurricane track, intensity, storm surge and to reduce forecast uncertainty, with an emphasis on rapid intensity change” (NOAA SAB 2007, p. 1). During July 2008–July 2009 the president’s proposed budget was amended to include \$13M for HFIP, and this increment became part of NOAA’s base budget.

This article describes the HFIP program, its goals, proposed methods for achieving those goals, and recent results from the program that suggest that it is on a path to meet its goals on time.

THE HURRICANE FORECAST IMPROVEMENT PROJECT. HFIP provides the unifying organizational infrastructure and funding for NOAA and other agencies to coordinate the hurricane research needed to significantly improve guidance for hurricane track, intensity, and storm surge forecasts. HFIP’s 5-yr (for 2014) and 10-yr (for 2019) goals are as follows:

- Reduce average track errors by 20% in 5 years and 50% in 10 years for days 1–5.
- Reduce average intensity errors by 20% in 5 years and 50% in 10 years for days 1–5.
- Increase the probability of detection (POD)¹ for rapid intensification (RI) change² to 90% at day 1 decreasing linearly to 60% at day 5, and decrease the false alarm ratio (FAR) for RI change to 10% for day 1 increasing linearly to 30% at day 5. The focus on RI change is the highest-priority forecast challenge identified by the National Hurricane Center (NHC).
- Extend the lead time for hurricane forecasts out to day 7 (with accuracy equivalent to that of the day-5 forecasts when they were introduced in 2003).

Forecasts of higher accuracy and greater reliability are expected to lead to higher user confidence and improved public response, resulting in savings of life and property. Reaching these goals, however, requires major investments in enhanced observational strategies, improved data assimilation, numerical model systems, expanded forecast applications based on the high-resolution and ensemble-based numerical prediction systems, and improved computational

infrastructure. NOAA also recognizes that addressing the challenges associated with improving hurricane forecasts requires interaction with, and the support of, the larger research and academic communities.

It is hypothesized that the ambitious HFIP goals could be met with high-resolution (~10–15 km) global atmospheric numerical forecast models run as an ensemble in combination with, and as a background for, regional models at even higher resolution (~1–5 km). In order to support the significant computational demands of such an approach, HFIP developed a high-performance computational system in Boulder, Colorado. Demonstrating the value of advanced science, new observations, higher-resolution models, and postprocessing applications is necessary to justify obtaining the commensurate resources required for robust real-time use in an operational environment.

For fiscal year (FY) 2011, HFIP program funding was approximately \$23M, with \$3M dedicated to enhancing computer capacity available to the program. The funding for computing was used to enhance the HFIP system established in Boulder, Colorado in FY2009, and resulted in machines called t-jet and u-jet with a total of 16,000 processors. The \$23M total includes ~\$7M of partial base funding for the NHC, Atlantic Oceanographic and Meteorological Laboratory (AOML)/Hurricane Research Division (HRD), Environmental Modeling Center (EMC) at the National Centers for Environmental Prediction (NCEP) and Earth System Research Laboratory (ESRL). The remaining \$13M was distributed to 1) various NOAA laboratories and centers, including the Geophysical Fluid Dynamics Laboratory (GFDL), National Environment Satellite, Data, and Information Service (NESDIS), ESRL, and NHC; 2) the National Center for Atmospheric Research (NCAR); 3) the Naval Research Laboratory (NRL) in Monterey; and 4) several universities—University of Wisconsin, The Pennsylvania State University (PSU), Colorado State University (CSU), The Florida State University (FSU), University of Wisconsin, and University of Rhode Island (URI; awarded through a NOAA announcement of opportunity)—and the National Oceanographic Partnership Program (NOPP). Specifically, \$1M was contributed each year for 3 years to the NOPP and, through an “Announcement

¹ The POD is equal to the total number of correct events forecast (hits) divided by the total number of events observed. The false alarm ratio (FAR) is equal to the total number of incorrect events forecast (misses) divided by the total number of events observed.

² RI for hurricanes is defined as an increase in wind speed of at least 30 knots (kt; 1 kt = 0.51 m s⁻¹) in 24 hours. This goal for HFIP also applies to rapid weakening of a decrease of 25 kt in 24 hours.

of Opportunity,” for completed proposals related to improving the understanding and prediction of hurricanes. The funding to NOPP from HFIP was matched by funding from the Office of Naval Research (ONR).

Distribution of the \$13M was based on recommendations from nine teams focused on various components of the hurricane forecast problem. The current teams, made up of over 50 members drawn from the hurricane research, development, and operational communities, are listed in Table 1 along with the team coleaders and the participating organizations.

HFIP’s focus and long-term goal is to improve the numerical model guidance that is provided by NCEP operations to NHC as part of the hurricane forecast process. To accomplish this goal, the program is structured along three parallel development paths, known as “streams.” Stream 1 is directed toward developments that can be accomplished using operational computing resources (either existing or planned). This stream covers development work planned, budgeted, and executed over the near term (mostly one to two years) by EMC with HFIP augmenting support to enable participation by the broader modeling community. Since Stream 1 enhancements are implemented into operational forecast systems, these advances are automatically

available to the hurricane specialists at NHC in the preparation of official forecast and warning products.

While Stream 1 works within presumed operational computing resource limitations, Stream 2 activities assume that resources will be found to greatly increase available computer power in operations above that planned for the next five years. The purpose of Stream 2 is to demonstrate that the application of advanced science, technology, and increased computing will lead to the desired increase in accuracy and other aspects of forecast performance. Because the level of computing necessary to perform such a demonstration is large, HFIP is developing its own computing system at NOAA/ESRL in Boulder, Colorado.

A major component of Stream 2 is an Experimental Forecast System (EFS) that HFIP runs each hurricane season. The purpose of the EFS (also known as the Demonstration Project) is to evaluate the strengths and weaknesses of promising new approaches that are demonstrable only with enhanced computing capabilities. The progress of Stream 2 work is evaluated each off-season to identify techniques that appear particularly promising to operational forecasters and/or modelers. These potential advances can be blended into the operational implementation plans through subsequent Stream 1 activities, or developed

TABLE 1. HFIP development teams.

Teams	Team leads and member’s organization
Global model/physics	<i>Stan Benjamin</i> (ESRL), <i>John Brown</i> (ESRL), AOML, NRL, GFDL, EMC, NRL
Regional model/physics	<i>Morris Bender</i> (GFDL), <i>Young Kwon</i> (EMC), AOML, NRL, ESRL URI, Old Dominion University, NCAR
Ensembles	<i>Zoltan Toth</i> (ESRL), <i>Carolyn Reynolds</i> (NRL), AOML, PSU, EMC, NHC, FSU
Data assimilation/vortex initialization team	<i>Jeff Whitaker</i> (ESRL), <i>Tomi Vukicevic</i> (AOML), NRL, Cooperative Institute for Research in the Atmosphere (CIARA), PSU, University of Oklahoma (OU)
Verification team	<i>Tim Marchok</i> (GFDL), <i>Barb Brown</i> (NCAR), NRL, NESDIS/Center for Satellite Applications and Research (STAR), AOML, NHC, EMC, ESRL, NWS/ Office of Science and Technology (OST)
Applications development and diagnostics	<i>Mark DeMaria</i> (NESDIS/STAR), <i>Ed Rappaport</i> (NHC), EMC, NRL, AOML, Research Applications Laboratory (RAL), ESRL, OU, AOML, FSU, NHC, AOML, National Weather Service/Office of Science and Technology (NWS/OST)
Hurricane observations	<i>Sim Aberson</i> (AOML), <i>John Knaff</i> (NESDIS/STAR), NHC, EMC, NESDIS/STAR, ESRL, URI, NRL, NOAA/Aircraft Operations Center (AOC), RAL, Rosenstiel School of Marine and Atmospheric Science (RSMAS), NCEP Central Operations (NCO), NCAR, NWS/OST
Coupled air–sea interaction	<i>Hyun-Sook Kim</i> (EMC), <i>George Halliwell</i> (AOML), URI, ESRL, NRL, RSMAS
Societal impacts	<i>Rick Knabb</i> (NHC), <i>Jennifer Sprague</i> [NWS/ Office of Assistant Administrator (OASST)], NWS/Southern Region (NWS/SR), NWS/Eastern Region (NWS/ER), Federal Emergency Management Agency (FEMA), Connecticut Emergency Management (CT-EM), Texas Emergency Management (TX-EM), North Carolina Emergency Management (NC-EM), Florida Emergency Management (FL-EM), Weather Channel

further outside of operations within Stream 2. Stream 2 models represent cutting-edge approaches that have little or no track record; consequently, NHC forecasters do not use these models to prepare their operational forecasts or warnings.

HFIP was originally structured around this two-stream approach. However, it quickly became apparent that some Stream 2 research models were producing forecast guidance that was potentially useful to forecasters. Because these models could not be implemented at NCEP because of insufficient operational computing resources, a third activity, known as Stream 1.5, was initiated to expedite the testing and availability of promising new models to forecasters. Stream 1.5 is a hybrid approach that accelerates the transfer of successful research from Stream 2 into real-time forecasting by following a path that temporarily bypasses the budgetary and technical bottlenecks associated with traditional operational implementations.

The Stream 1.5 process for each upcoming hurricane season involves extensive evaluation of the previous season's most promising Stream 2 models or techniques. This testing involves rerunning the models or techniques over storms selected by NHC from several previous seasons involving several hundred cases. For those that meet certain predefined standards for improvement over existing techniques, and if operational computing resources are not available for immediate implementation, these enhancements can be run on HFIP computing resources and be provided to NHC forecasters in real time during the upcoming hurricane season as part of the EFS. This process moves forward the availability of real-time advances to forecasters one or more years. It also serves as a

proof of concept for both the developmental work (Stream 2) and augmented computational capabilities.

THE HFIP MODEL SYSTEMS. HFIP believes that the best approach to improving hurricane track forecasts, particularly beyond four days, involves the use of high-resolution global models, with at least some run as an ensemble. However, global model ensembles are likely to be limited by computing capability for at least the next five years to a resolution no finer than about 15–20 km, which is inadequate to resolve the inner core of a hurricane. The HIRWG asserted that the inner core must be resolved to expect to see consistently accurate hurricane intensity forecasts (NOAA SAB 2006). Maximizing improvements in hurricane intensity forecasts will therefore likely require high-resolution regional models, perhaps also run as an ensemble. Below we outline the modeling systems currently in use by HFIP.

The global models. Global models provide the foundation for all of HFIP's modeling effort. They provide hurricane forecasts of their own, and are top-tier performers for hurricane track. They also provide background data and/or boundary conditions for regional and statistical models and can be used to construct single-model ensembles, or be members of multimodel ensembles. The HFIP EFS involves three global models: ESRL's Flow-Following Finite-Volume Icosahedral Model (FIM), NCEP's Global Forecast System (GFS), and NRL's Navy Operational Global Atmospheric Prediction System (NOGAPS).

The FIM is an experimental global model that can be run at various resolutions and uses initial conditions from a number of sources (Benjamin et al.

TABLE 2. Specifications of the HFIP global models. LSM = land surface model. RRTM = rapid radiative transfer model.

Models	Horizontal resolution (km)	Vertical levels	Cumulus parameterization	Microphysics	PBL	Land surface	Radiation	Initialization
FIM	27	64	From 2010GFS—simplified Arakawa–Schubert	Ferrier	GFS nonlocal PBL	Noah LSM	GFDL/RRTM	ESRL EnKF
GFS/EnKF	27	64	Simplified Arakawa–Schubert	Ferrier	GFS nonlocal PBL	Noah LSM	GFDL scheme	ESRL EnKF
GFS/GSI	27	64	Simplified Arakawa–Schubert	Ferrier	GFS nonlocal PBL	Noah LSM	GFDL	GSI
NOGAPS	41	42	Emanuel	—	NOGAPS	NOGAPS	Harshvardhan/Fu–Liou	NAVDAS-AR

2004; Bao et al. 2012). It is currently using a constant sea surface temperature underneath.

The GFS, the NWS's global model, currently has two versions in use by the HFIP EFS. One of these is the current operational model run at NOAA and NCEP. The second is an experimental version developed at ESRL, which differs from the operational GFS by featuring a fixed ocean and an ensemble Kalman filter (EnKF)-based initialization system (see the section "Initialization and data assimilation systems").

HFIP currently is using the operational NOGAPS model. A semi-Lagrangian version is being developed, which will allow for efficient high-resolution forecasts (NOARL 1992).

Some specifics of the global models are shown in Table 2.

The regional models. Specifics of the regional models are shown in Table 3. Note that GFDL (OPS) and Hurricane Weather Research and Forecasting (HWRF) models (OPS) refer to the current operational regional models. The Weather Research and Forecasting (WRF) modeling system in use by HFIP contains two options for its dynamic core, and several options for physics as well as initialization and postprocessing systems (DTC 2011b). The two dynamic core configurations are the Weather Research and Forecast model (ARW-WRF), built

TABLE 3. Specifications for the HFIP regional models. UW NMS = University of Wisconsin Nonhydrostatic Modeling System. YSU = Yonsei University.

Models	Nesting/ horizontal resolution (km)	Vertical levels core	Cumulus parameterization	Microphysics	PBL	Land surface	Radiation	Initial and boundary conditions	Initialization
HWRF (OPS)	2 27/9	42 NMM	Simplified Arakawa-Schubert	Ferrier	GFS nonlocal PBL	GFDL slab model	GFDL scheme	GFS	GSI 3DVAR
GFDL (OPS)	3: 30/15/7.5	42 GFDL	Arakawa-Schubert	Ferrier	GFS nonlocal PBL	Slab model	Schwarzkopf-Fels scheme	GFS	GFDL synthetic bogus vortex
HWRF-HRD/ EMC 27-9-3	3: 27/9/3	42 NMM	Simplified Arakawa-Schubert	Ferrier	GFS nonlocal PBL	GFDL slab model	GFDL scheme	GFS	EnKF with aircraft data
WRF ARW (NCAR)	3: 12/4/1.3	36 ARW	New Kain-Fritsch (12 km only)	WSM5	YSU	Five-layer thermal diffusion soil model	RRTM (longwave)/ Dudhia (shortwave)	GFS	EnKF method in a 6-h cycling mode
COAMPS-TC	3: 45/15/5 (15/5 km following the storm)	40 COAMPS	Kain-Fritsch on 45- and 15-km meshes	Explicit microphysics (five-class bulk scheme)	Navy 1.5 order closure	Slab with the NOAH LSM as an option	Harshvardhan	NOGAPS (could use GFS if desirable)	3DVAR data assimilation with synthetic observations
Wisconsin model	UW NMS (3D entropy/ entropy/ KE conserving dynamics core)	3-4: 90/45/9 km 90/45/9/3 km	52	Modified Emanuel	Explicit bulk microphysics (cloud/rain/pristine/ aggregate/graupeil)	1.5-order closure	WRF vegetation/ land surface/ Andreas emulsion layer	RRTM	GFS/GFDL
PSU ARW	3: 40.5/13.5/4.5 for ensemble forecast 1.5-km nest for control	35 ARW	Grell-Devenyi ensemble scheme (40.5 km only)	WSM six-class graupeil scheme	YSU	Five-layer thermal diffusion scheme	RRTM (longwave)/ Dudhia (shortwave)	GFS	EnKF with NOAA airborne radar

by NCAR, and the Nonhydrostatic Mesoscale Model (NMM), built by EMC.

The operational NCEP HWRF derives from the NMM dynamic core and has a movable, two-way nested grid capability with an inner nest that covers a $6^\circ \times 6^\circ$ region at 9-km resolution. A coarser outer domain covers a $75^\circ \times 75^\circ$ region at 27-km resolution. The model has 42 vertical layers. Advanced physics include atmosphere/ocean fluxes, coupling with the Princeton Ocean Model, and the NCEP GFS physics. The 2012 operational HWRF, developed by EMC and AOML with HFIP support, added a third nest covering a $6^\circ \times 5^\circ$ region at 3-km horizontal resolution within the second moving nest that now covers an $11^\circ \times 10^\circ$ region at 9-km resolution.

HFIP also supports the WRF ARW system, which NCAR runs using a simplified one-dimensional model of the ocean. It features three interactive nests with an inner-nest resolution of 1.3 km.

The Pennsylvania State University (PSU) regional ensemble constitutes another version of the WRF ARW system, with similarities to the NCAR WRF ARW. It uses a static interactive inner nest of 4 km but no interactive ocean (PSU 2011; Zhang et al. 2011; Weng and Zhang 2012; Snyder and Zhang 2003).

The Coupled Ocean–Atmosphere Mesoscale Prediction System Tropical Cyclone (COAMPS-TC) and the Wisconsin model (WISC 2011) are detailed in the table and have been members of the Stream 1.5 suite of models each year. Note that COAMPS-TC features an interactive ocean (NRL 2011).

High-resolution ensemble approach. A single forecast from a particular numerical model (often referred to as a “deterministic” run) has an inherent level of uncertainty. An ensemble or collection of forecasts all valid for the same time, however, can potentially provide information on the amount of confidence that can be associated with that forecast situation. In addition, the mean forecast of a well-constructed ensemble is often superior to the forecast from any individual member of the ensemble. Ensembles, therefore, offer the potential to improve both forecast accuracy and forecast utility.

High resolution is necessary in these ensembles in order to adequately resolve the hurricane structure (NOAA SAB 2007), because the hurricane can alter the flow in which it is embedded and, in turn, this altered flow will impact the hurricane track and potentially also its intensity. To even begin to get structures in forecast models that resemble actual hurricanes, resolutions of 15–20 km are likely necessary and 1–5-km resolution will be necessary to adequately re-

solve the inner core structure. Ideally, each ensemble member would have the same resolution, and at least 20–30 members will need to be computed to provide adequate estimates of the uncertainty.

It has been shown that the evolution of the atmospheric flow at a given location beyond about three days depends on atmospheric features distributed globally (Reynolds et al. 2009; Hakim 2003; Langland et al. 2002; Palmer et al. 1998; Rabier et al. 1996; Hoskins and Ambrizzi 1993; Chang 1993; and others). It is therefore not surprising that regional models are generally outperformed by the global models for hurricane track forecasts beyond three days, a degradation likely related to the need for specifying lateral boundary conditions and the inward propagation of errors from that specification affecting the hurricane track. HFIP believes that tropical cyclone forecasts that extend out to 5–7 days will require the forecast models be global in nature and, for the reasons noted above, be run at the highest resolution possible. Fortunately, on the experimental computing established by HFIP, it is now possible for deterministic global models to be run at nearly the same resolution as the current regional models (10–15 vs 3–4 km).

Although there are many ways to create an ensemble for tropical cyclone prediction, the most successful ensembles to date have been multimodel ensembles, such as the “TVCN” aid used in operations at NHC (Cangialosi and Franklin 2011; see also NHC 2011). TVCN is simply the mean of the forecast tracks from the GFS, the Met Office global model (UKMET), NOGAPS, HWRF, GFDL, the U.S. Navy’s version of the GFDL (GFDN), and the European Centre for Medium-Range Weather Forecasts (ECMWF) model. The spread of the forecasts making up this particular ensemble has also been used operationally at NHC to assess forecast uncertainty (Goerss 2007). A similar multimodel ensemble is used operationally for intensity prediction; intensity consensus (ICON) is the mean of the forecasts from NHC’s top-performing operational intensity models—Decay–Statistical Hurricane Intensity Prediction System (SHIPS), the Logistics Growth Equation Model (LGEM; NHC 2011), GFDL, and HWRF. Forecasters often refer to the mean forecast from a multimodel ensemble as a consensus.

Ensembles can also be formed using a single dynamical model, either by slightly varying the model’s initial condition or altering some component of the model, such as the physics package or a model parameter. Initially small differences among the ensemble members will grow with time, at rates that depend on the weather situation. Frequently, but not always,

the highest probability is that the correct forecast is near the mean, median, or mode of the ensemble, although other ensemble realizations have a finite probability of being correct (Buizza 1997), and of course the actual track can lie outside the envelope of ensemble member tracks.

As computing power and model resolution increase, the accuracy of the single-model ensemble mean should improve. But perhaps more promising is the interpretive and diagnostic potential of the ensemble approach. For forecasters, exploring the reasons behind the variability in a set of ensemble forecasts, such as the interplay between intensity and track, can help the forecaster determine an outcome more likely than just the ensemble mean. Much the same can be said for regional ensembles, but here the emphasis shifts from longer-range forecasts of track to medium-range forecasts of intensity. Since intensity change is thought to reside largely in the dynamics of the inner core region of the tropical cyclone, the inner core must be resolved to scales of 1–5 km.

Initialization and data assimilation systems. A number of approaches are used to create the initial state for the global and regional models in the HFIP EFS.

1) GLOBAL FORECAST SYSTEM

The initial state created for the current operational global model (GFS) is interpolated to the grids used by HFIP global models. The GFS uses the Gridpoint Statistical Interpolation (GSI) initialization system that has been run operationally since 2006 and is a three-dimensional variational data assimilation (3DVAR) approach (DTC 2011a; Purser et al. 2003a,b; Wu et al. 2002; Parrish and Derber 1992; Cohn and Parrish 1991).

2) HWRF

The operational HWRF uses an advanced vortex initialization and assimilation cycle consisting of four major steps: 1) interpolation of the global analysis fields from the GFS onto the operational HWRF model grid; 2) removal of the GFS vortex from the global analysis; 3) addition of the HWRF vortex modified from the previous cycle's 6-h forecast (or use of a bogus vortex for a cold start); and 4) addition of satellite radiance and other observation data in the hurricane area (9-km inner domain) using GSI.

3) NRL ATMOSPHERIC VARIATIONAL DATA ASSIMILATION SYSTEM (NAVDAS)

This is the system used to provide the initial conditions to NOGAPS. Previously a 3DVAR

system, it was upgraded in September 2009 to NAVDAS-AR, a four-dimensional variational (3DVAR) approach (NRL 2011; Daley and Barker 2001). The 3DVAR version of NAVDAS is used to initialize COAMPS-TC.

4) ENSEMBLE KALMAN FILTER

This is an advanced assimilation approach, somewhat like 3DVAR, that uses an ensemble to create background error statistics for a Kalman filter (Tippett et al. 2003; Keppenne 2000; Evensen 1994; Houtekamer and Mitchell 1998). While this approach is still in the experimental stage in the United States, it has shown considerable promise (Hamill et al. 2011).

5) HURRICANE ENSEMBLE DATA ASSIMILATION SYSTEM (HEDAS)

HEDAS is an EnKF system applied to the HWRF and was developed at AOML (Aksoy et al. 2012).

6) HYBRID VARIATIONAL-ENSEMBLE DATA ASSIMILATION SYSTEM (HVEDAS)

This system combines aspects of the EnKF and 3D- or 4DVAR, such as using the ensemble of forecasts to estimate the covariances at the start of the variational component of the DA system. This technology was developed at EMC, ESRL, and AOML and was used in operations for the 2012 season. This is commonly referred to as the hybrid data assimilation system.

7) BOGUS VORTEX

The initial state for some of the regional models is produced by removing the vortex from the first guess and then inserting a new vortex; when the new vortex is defined, rather than analyzed, it is known as a “bogus” vortex. This is similar to the HWRF initialization system but for HWRF the introduction of the bogus vortex was followed by a GSI data assimilation (DA) cycle. Bogus vortex relocation is used by the current operational Global Ensemble Forecast System (GEFS) run by NCEP. None of the HFIP global models is currently using vortex relocation.

The HWRF community code repository. During 2009–11, both EMC and the Developmental Testbed Center (DTC) worked to update the operational version of HWRF from version 2.0 to the current version of HWRF at the DTC (version 3.2; DTC 2011b). This makes the operational model completely compatible with the codes in the central DTC repository, allows researchers access to the operational codes, and makes improvements in HWRF developed by the research community easily transferable into operations. This was one of the initial goals of the WRF program.

MEETING THE HFIP GOALS. *The HFIP baseline.* To measure progress toward meeting the HFIP goals outlined in the introduction, a baseline level of accuracy was established to represent the state of the science at the beginning of the program. Results from

HFIP model guidance could then be compared with the baseline to assess progress. HFIP accepted a set of baseline track and intensity errors developed by NHC, in which the baseline was the consensus (average) from an ensemble of top-performing operational models,

evaluated over the period 2006–08. For track, the ensemble members were the operational aids GFS model (GFSI), GFDL model (GFDI), UK model (UKMI), NOGAPS model (NGPI), HWRF model (HWFI), the U.S. Navy's operational version of the GFDL model (GFNI), and ECMWF model (EMXI), while for intensity the members were the operational GFDL model (GHMI), the operational HWRF (HWFI), the Decay Statistical Hurricane Intensity Prediction Scheme model (DSHP), and the SHIPS Logistic Growth Equation Model (LGEM) (Cangialosi and Franklin 2011). Figure 1 shows the mean errors of the consensus (CONS) over the period 2006–08 for the Atlantic basin. A separate set of baseline errors (not shown) was computed for the eastern North Pacific basin.

The baseline errors in Fig. 1 are also compared to the errors for the same cases for the climatology and persistence (CLIPER) models (track; we use CLIPER5) and Decay-Statistical Hurricane Intensity Forecast, version 5 (SHIFOR5) for intensity (NHC 2011). Errors from these two models are large when a storm behaves in an unusual or rapidly changing way, and therefore are useful in assessing the inherent difficulty in a set of forecasts. When a track or intensity model error is normalized by the CLIPER5 or Decay-SHIFOR5 error, the normalization yields a measure of the model's skill.

Since a sample of cases from, say, the 2011 season might have a different inherent level of difficulty from the baseline sample of 2006–08 (e.g., because it had an unusually high or low number of rapidly intensifying storms), it is necessary to evaluate the progress of the HFIP models

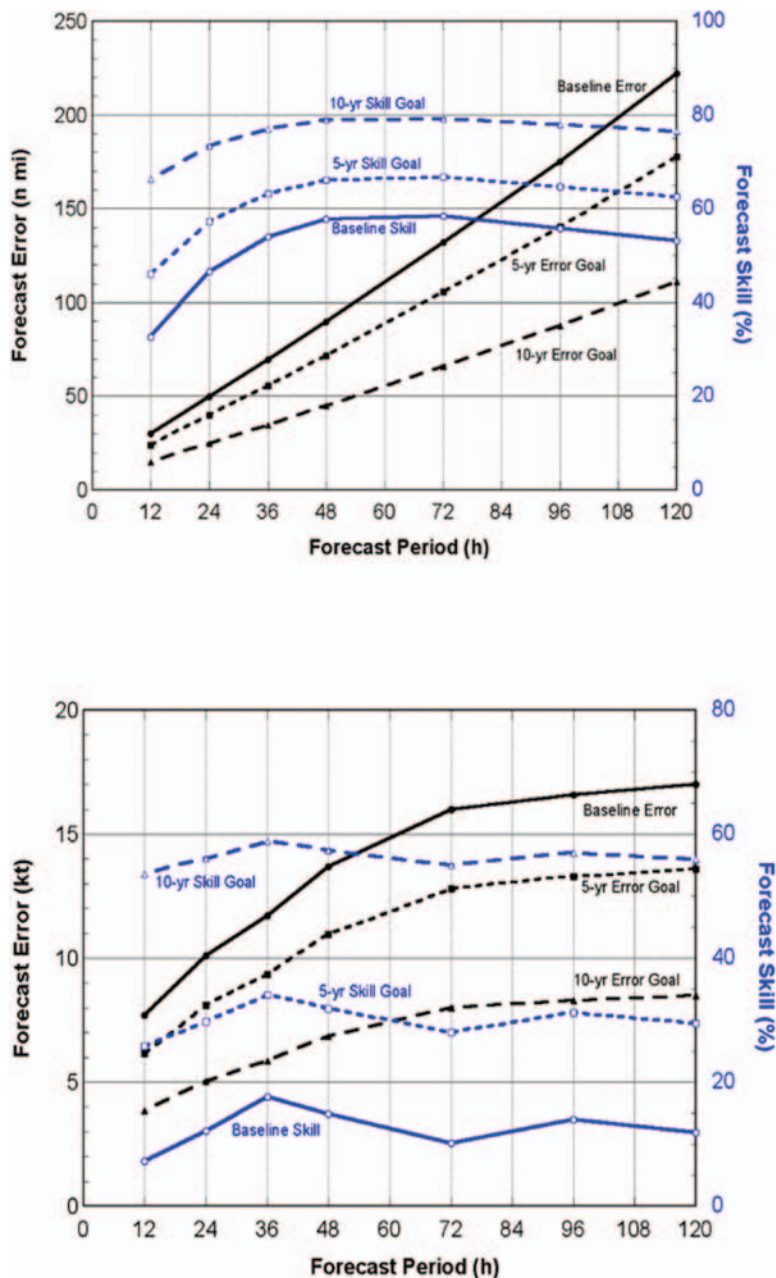


FIG. 1. HFIP (top) baseline track and (bottom) intensity errors. The baseline errors (solid black lines) were determined from an average of the top-flight operational models during the period 2006–08. The HFIP expressed goals (black dashed lines) are to reduce these errors by 20% in 5 years and by 50% within 10 years. In order to permit comparisons of nonhomogeneous samples, the baseline errors and HFIP goals are also expressed in terms of skill relative to a climatology and persistence standard (see text). The skill baselines and goals are shown by the solid and dashed blue lines, respectively.

in terms of forecast skill, rather than in terms of error. Figure 1 displays the skill of the baseline errors and the 5- and 10-yr goals represented in blue, and labeled on the right side of the graph is the percentage improvement over the SHIFOR5 and CLIPER5 forecasts for the same cases. Note the baseline skill for intensity at all lead times is roughly constant with the baseline, representing a 10% improvement over SHIFOR5 and the 5- and 10-yr goals, 30% and 55%, respectively.

It is also important to note that these HFIP performance goal baselines were determined from a class of operational aids known as “early” models. Early models are those that are available to forecasters early enough to meet forecast deadlines for the synoptic cycle. Nearly all the dynamical models currently in use at tropical cyclone forecast centers, however (such as the GFS or GFDL), are considered “late” models because they arrive too late to be used in the forecast for the current synoptic cycle. For example, the 1200 UTC GFDL run does not become available to forecasters until around 1600 UTC, whereas the NHC official forecast based on the 1200 UTC initialization must be issued by 1500 UTC, 1 hour before the GFDL forecast can be viewed. It is actually the older (0600 UTC) run of the GFDL that would be used as input for the 1500 UTC official NHC forecast, through a procedure was developed to adjust the 0600 UTC model run to match the actual storm location and intensity at 1200 UTC. This adjustment procedure creates the 1200 UTC “early” aid GFDI that can be used for the 1500 UTC NHC forecast. The distinction between early and late models is important to assessing model performance, since late models have an advantage of more recent observations/analysis than their early counterparts.

Meeting the track goals. Earlier we noted that accurate forecasts beyond a few days require a global domain because influences on a forecast for a particular

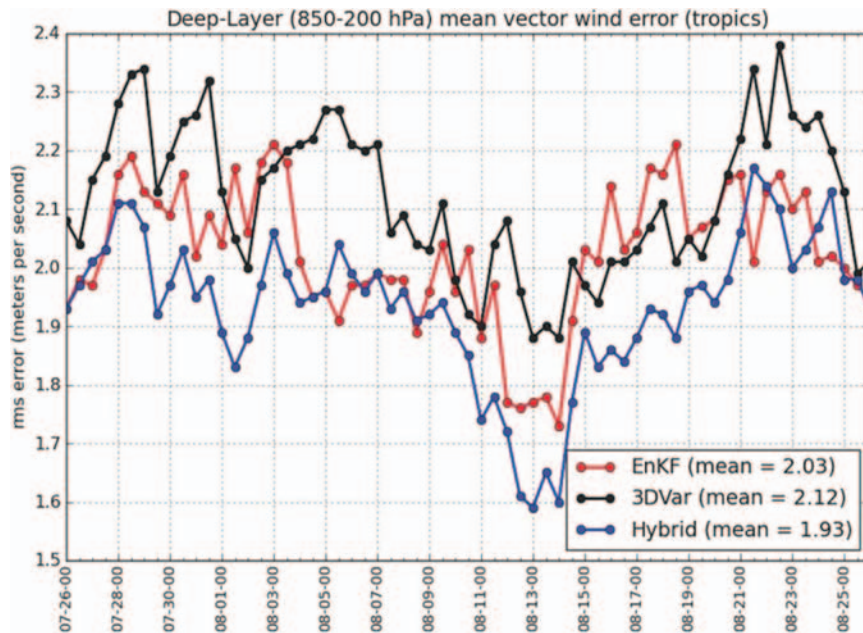


FIG. 2. Verification statistics (RMS error) for 72-h forecasts of the deep-layer-mean wind in the tropics, run using the GFS operational model at 30-km resolution with initial conditions specified by three different initializations: the GSI 3DVAR operational data assimilation system, the experimental EnKF data assimilation system, and the experimental hybrid data assimilation system. The number in parentheses in the legend shows the mean RMS vector wind error for each configuration over the evaluation sample. Date is indicated along the horizontal axis in MM-DD-YY format at 0000 UTC. (J. Whitaker and D. Kleist 2011, personal communication).

location come from weather systems at increasing distance from the local region over time. One of the first efforts in HFIP, therefore, was to improve the existing operational global models. Early in the program, it was shown that using a more advanced data assimilation scheme than the one currently employed operationally at NCEP (GSI) improved forecasts, particularly in the tropics.

Figure 2 compares forecasts of tropical winds in the GFS using the GSI, ensemble Kalman filter, and a hybrid approach that is a combination of the GSI and EnKF (see the section “Initialization and data assimilation systems”). Both the hybrid and the EnKF data assimilation approaches outperform the GSI initialization. All HFIP global models now use the EnKF system, and most of the regional models will eventually adopt the hybrid system as well. Based on comparisons such as these, NCEP replaced the GSI data assimilation system with a hybrid system (Hamill et al. 2011) in May 2012. Note that the hybrid system performs better than the EnKF system alone.

Figure 3 shows the tropical cyclone track forecast skill (see the section “The HFIP baseline”) for various regional and global models, including the operational

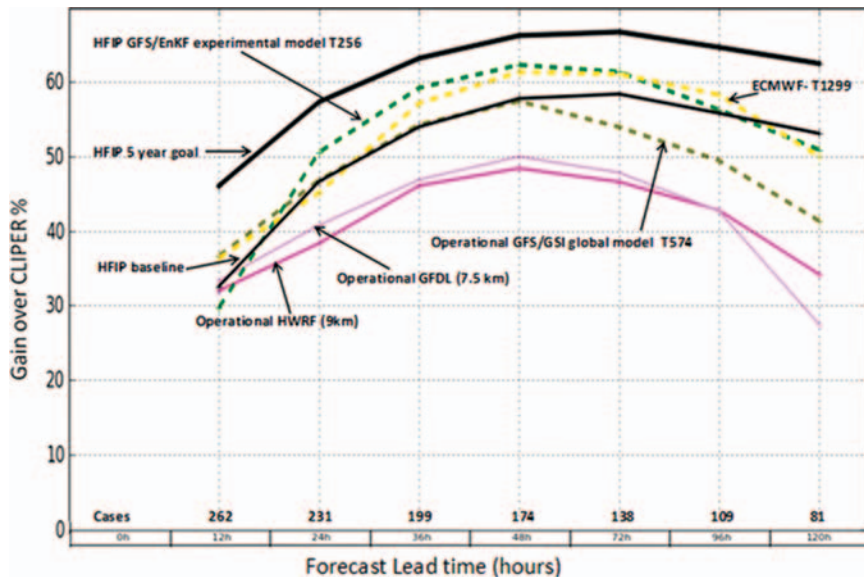


FIG. 3. Comparison of selected model track skill with the HFIP baseline and 5-yr goal, evaluated over the period 2010–11. The baseline skill and the 5-yr HFIP intensity goal (see Fig. 2) are shown in black. The number of cases at each forecast lead time is shown above the x axis. (M. Fiorino 2012, personal communication.)

deterministic global models GFS GSI (3DVAR initialization) and ECMWF (with a 3DVAR initialization) and the two operational regional hurricane models, HWRf and GFDL, for the 2010–11 hurricane seasons. All the models shown in Fig. 3 are late models (see the section “The HFIP baseline”). The GSF/ENKF T256 model shown in the figure is the HFIP ensemble with 20 members (ensemble mean shown) run at low resolution, T254 (~60 km), while the operational deterministic GFS was run at T574 (~30 km), and the ECMWF was run at T1299 (~15 km). Note that even though the GFS/EnKF ensemble has the lowest resolution of any of the models shown, it still outperformed the other guidance. In fact, the GFS ensemble is close to the HFIP 5-yr goal of a 20% improvement over the baseline for almost half of the forecast lead times. Note also that the track forecast skill from the regional models is less than that of the global models, a behavior typical of most regional models. The superior performance of the GFS with EnKF ensemble is at least partially due to the data assimilation system, since the model is exactly the same as the operational GFS/GSI and the ensemble was run at lower resolution than the GFS/GSI (T574, ~30 km).

Reaching the intensity goals. HFIP expects that its intensity goals will be achieved through the use

of regional models with a horizontal resolution near the core finer than about 3 km. In addition, early results suggest that output from individual HFIP models can be used in statistical models such as SHIPS (DeMaria and Kaplan 1994; NHC 2011) or LGEM (DeMaria 2009; NHC 2011) to further increase intensity forecast skill.

The suite of models that comprised the 2011 HFIP regional model ensemble is listed in Table 3. Components of some of these models qualified for Stream 1.5 (see HFIP 2012 for details). Figure 4 shows a homogeneous sample of

operational and Stream 1.5 models plotted during the 2011 hurricane season as skill relative to SHIFOR5. Note that COAMPS-TC and the SPC3 statistical model (both HFIP models) showed a 5%–10% improvement over SHIFOR5 during the early part of the forecast, but beyond 72 hours all guidance dropped off in skill, though the number of cases at 120 hours is very small. Most models have little or no skill compared to SHIFOR5, which is why there has been little improvement in intensity forecasts.

Two research groups—one at HRD led by Altug Aksoy (using an experimental version of the HWRf model with a double nest at 9- and 3-km resolution and HEDAS (Aksoy et al. 2012)³ and the other headed by Fuqing Zhang at the Pennsylvania State University (using the ARW model at 4-km resolution)—demonstrated that assimilating data collected by the NOAA P-3 tail Doppler radar as well as other data collected by the P-3s, the NOAA G-IV, and Air Force C-130 hurricane hunter aircraft (including flight-level and dropsonde data) with an EnKF system can improve intensity forecasts. The results compared to SHIFOR5 are shown in Fig. 5 for cases for which radar was obtained from 2008 to 2011. HEDAS (solid green line) assimilated all available aircraft data, whereas the PSU system (dashed green line) assimilated only the tail Doppler radar data. The red line

³ This version of HWRf is not the same as the operational HWRf shown in Fig. 4. The operational HWRf had an inner nest of 9 km, while the version of HWRf used in these calculations had an inner nest resolution of 3 km.

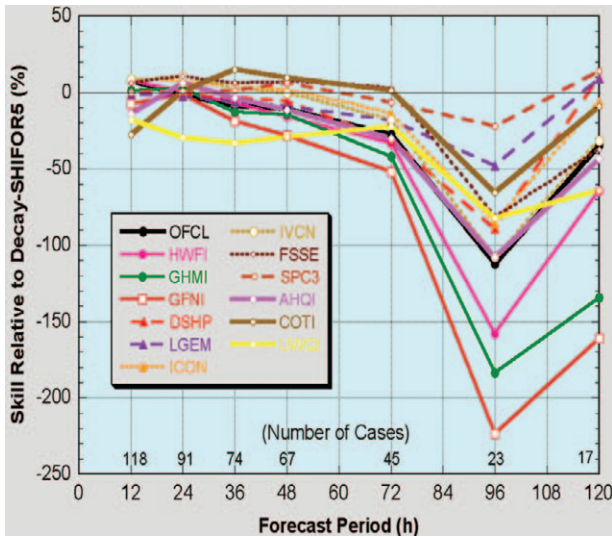


FIG. 4. Intensity forecast skill for a homogeneous sample of selected operational and HFIP Stream 1.5 models for the 2011 season. The dashed lines indicate statistical models, and the solid lines are the dynamical models. Forecast sources are OFCL = NHC official forecast; DSHP, LGEM, GHMI, HWFI; SPC3 = intensity consensus: six-member ensemble of DSHP and LGEM (three each) with predictors from GFS/GFDL/HWRF, respectively, for each member; GFNI; AHQI = NCAR model; COTI = COAMPS-TC with GFS initial conditions; UWQI = University of Wisconsin model; and FSSE = the Florida State Superensemble including the HFIP regional models list in Table 3. IVCN is the NHC intensity consensus (NHC 2011). Note the very small sample size at 96 and 120 hours.

shows HWRF with the standard initialization. The black lines (solid, dashed, and dotted) show the baseline skill and the 20% and 50% improvement goals, respectively. The red line can be compared with the solid green line to assess the impact of assimilating the aircraft data.

From 36 to 96 hours, the high-resolution models improve forecasts by as much as 50% over SHIFOR5, close to or exceeding the 20% HFIP goal and approaching the 50% goal for these cases. Much of this improvement comes from the model itself, perhaps from the higher resolution used as compared to the operational HWRF (compare in Fig. 5 the solid red and green lines to the zero skill line). But assimilating aircraft data (compare the solid red and green lines) provides a further 10% improvement in skill and these improvements are statistically significant at the 95% level at 24, 36, and 60 hours. The size of the errors from the PSU system is consistent with those from the HEDAS runs, lending confidence to this result. This is encouraging, given that the average skill of the current operational dynamical models (GFDL and HWRF) is 20% or less (Cangialosi and Franklin 2011). The experimental models perform as well as or statistically significantly better than the operational intensity consensus (not shown) for this sample. Solving the problem in the early part of the forecast when the model is adjusting to the initial conditions is a major current focus of HFIP.

Statistical postprocessing of model output. Much of the discussion above focused on using numerical model improvements to achieve the HFIP goals. Typically statistical models (e.g., DSHP and LGEM) perform among the best as predictors of hurricane intensity. A statistical model is one where a limited number of

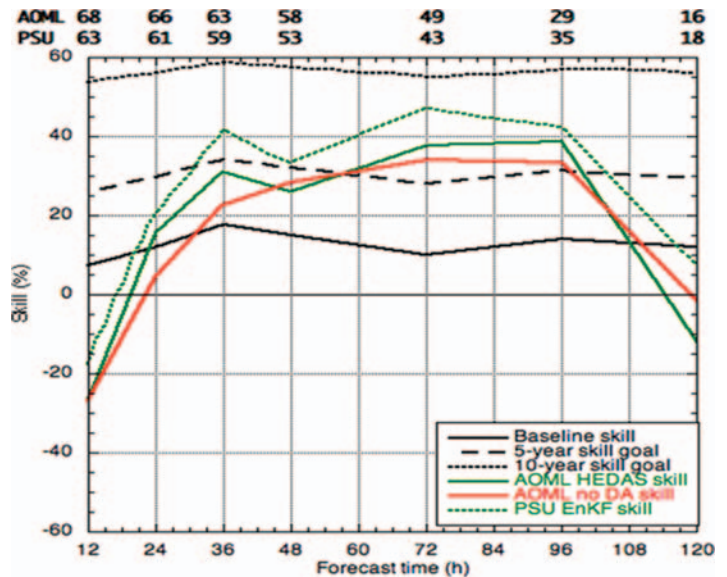


FIG. 5. Intensity forecast skill from two HFIP models that can assimilate aircraft data from the tropical cyclone core, for a sample of cases from 2008 to 2011. The solid green line shows the mean skill for the HWRF using the AOML HEDAS data assimilation system to incorporate Doppler radar, aircraft flight-level, and dropwindsonde observations, while the solid red line shows the skill of the same model using the standard HWRF initialization without any of the inner core observations. The dashed green line represents the skill of the PSU ARW model with the Doppler data included. The HFIP skill baseline and goals are shown in black. Sample sizes for the AOML and PSU samples are given along the top of the diagram. [S. Abernethy (HRD) and F. Zhang (PSU) 2011, personal communication.]

predictors (measured in single to double digits) are combined with weights that are determined by correlation with past data. These predictors are generally selected from parameters describing the current state of the hurricane or various environmental data. Those using environmental data can specify their values from current observations of model forecasts. There is another class of statistical model that takes a particular prediction from a dynamical model (say, track or intensity) and combines it with a weighted average from other models in a multimodel ensemble. The weights are determined by comparing the performance of the various models over a period of years. Perhaps the simplest statistical model for intensity is SHIFOR5 in which the predictors are current position and intensity, the previous 12 hour position and intensity, and Julian date (CLIPER5 is a similar model for track). It is sobering that even a model this simple provides forecasts of intensity almost as good as any of the current dynamical models. More complex statistical models used operationally for intensity are SHIPS and LGEM (NHC 2011). SPC3 (results shown in Fig. 4) provides an improvement compared to the operational statistical and dynamical models by using multiple operational numerical models as input for the environmental predictors. In Fig. 4, SPC3 uses input from the operational GFS, HWRF, and GFDL models for both DSHP and LGEM. This gives six variations that are then averaged as an ensemble. In Fig. 4 SPC3 is among the best performers at all lead times so the use of the statistical models with data from dynamical models can be used to improve the predictions of the dynamical models. The SPC3 was also as good as or better than the other statisti-

cal models (LGEM and DSHP) so using the model forecasts in the statistical models improves them as well. Figure 6 provides further indications that dynamical model forecasts can be improved using statistical models. It shows a comparison between the current operational HWRF and GFDL models by themselves and a combination of DSHP/LGEM using parameters determined from each of those two operational models.

Note that when using both the GFDL and HWRF operational models for input, the statistical models gave an improvement of up to 20% (5 kt) over the parent operational model.

THE CONFIGURATION OF A NUMERICAL MODEL HURRICANE FORECAST GUIDANCE SYSTEM TO MEET THE HFIP GOALS.

While it appears that use of aircraft data will likely help HFIP meet its intensity goals for storms for which such data are available, these data will not be available for storms for a large majority of model initializations. For those we will need to rely on better use of satellite data taken in the near vicinity of the hurricane. A longer-term major focus for HFIP is to improve satellite data assimilation in regional model initialization systems.

In this article we have not addressed the goal of HFIP to improve the forecasting of rapid changes in tropical cyclone intensity because, at this juncture, none of the HFIP dynamical models is capable of providing reliable forecasts of rapid intensification. The global models are not able to resolve the inner core processes that are likely to be very important in the RI process, and all the regional models have serious spinup (and spindown) problems

(Fig. 5). Except for the RI issue, we can now say with considerable confidence what a final end-state operational configuration of the hurricane numerical prediction system should look like in 2014, the end of the initial 5 years of HFIP.

The longer-range predictions, out to 1 week, of both track and intensity will be accomplished by global models run as ensembles and initialized with a hybrid data assimilation system and postprocessed with various statistical models. Resolution of these global models needs to be no coarser than about 20 km, and the results will be improved if more than one global model is used in the ensemble.

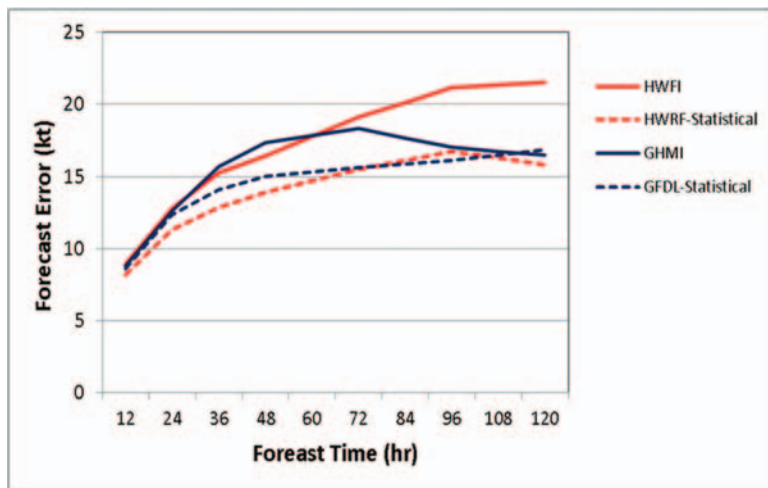


FIG. 6. Comparison of intensity forecast error for two statistical models and their parent dynamical models, for all tropical cyclones for the period 2008–11 (M. DeMaria 2011, personal communication.)

TABLE 4. Numerical model hurricane forecast guidance system.

Component	Specifications
Global model ensemble with hybrid data assimilation	· 20 members at 20 km
	· Multimodel (at least two; e.g., FIM, GSF)
Regional model ensemble	· 20 members at 3 km
	· Multimodel (at least two; e.g., HWRF, AHW, COAMPS-TC)
	· Using all available aircraft and satellite data in core and near environment of hurricane
Statistical postprocessing	· LGEM, SHIPS, others

The intensity goals for forecast periods out to 48–72 hours will be accomplished with regional models run with resolution at least as fine as 3 km as a multimodel ensemble. All models will use all available aircraft and satellite data. These will also be postprocessed with statistical models. The focus with the regional models will be on intensity, and with the high resolution the RI goals may be met with the regional models. More specifically, the end system might include a global model ensemble with hybrid data assimilation, a regional model ensemble with hybrid data assimilation, and statistical postprocessing (Table 4). The ability to run this system, however, will require at least a tenfold increase in computer resources in operations in order to run the high-resolution ensembles.

CONCLUDING REMARKS. By reaching out across the hurricane research, development, and operational communities, HFIP promoted the cooperative effort necessary to make rapid improvements in hurricane forecast guidance. The focus on improving the data assimilation system in the global models and the use of ensembles from the global models are likely to lead to substantial improvements in hurricane track forecasts in operations in the near future. Use of high-resolution regional models with advanced data assimilation systems such as the hybrid system being developed at EMC, together with the use of aircraft and satellite data in the inner core area at the scale of the central hurricane region, will likely lead to improved forecasts of intensity. In both intensity and track, HFIP expects to reach its interim 5-yr goals of 20% improvement over the operational baselines that reflect the level of forecast skill at the start of the program.

ACKNOWLEDGMENTS. The authors acknowledge the entire HFIP community including all members of the HFIP teams (leads shown in Table 1; team members are

listed online at www.hfip.org/teams/), all members of the various modeling teams for the models listed in Tables 2 and 3, and the several members of the team, noted in the text, who contributed figures for the article.

REFERENCES

- Aksoy, A., S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Abernethy, and F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for high-resolution data: The impact of airborne Doppler radar observations in an OSSE. *Mon. Wea. Rev.*, **140**, 1843–1862.
- Bao, J.-W., S. Benjamin, R. Bleck, J. Brown, J. Lee, A. MacDonald, J. Middlecoff, and N. Wang, cited 2012: FIM documentation. NOAA, 20 pp. [Available online at http://fim.noaa.gov/fimdocu_rb.pdf.]
- Benjamin, S., G. Grell, J. Brown, T. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494.
- Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **125**, 99–119.
- , P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097.
- Cangialosi, J. P., and J. L. Franklin, 2011: 2010 National Hurricane Center forecast verification report. NOAA, 77 pp. [Available online at www.nhc.noaa.gov/verification/pdfs/Verification_2010.pdf.]
- Chang, E. K. M., 1993: Downstream development of baroclinic waves as inferred from regression analysis. *J. Atmos. Sci.*, **50**, 2038–2053.
- Cohn, S. E., and D. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**, 1757–1785.

- Daley, R., and E. Barker, 2001: NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.*, **129**, 869–883.
- DeMaria, M., 2009: A simplified dynamical system for tropical cyclone intensity prediction. *Mon. Wea. Rev.*, **137**, 68–82.
- , and J. Kaplan, 1994: A Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220.
- DTC, cited 2011a: Community Gridpoint Statistical Interpolation (GSI): Documents and publications. Developmental Testbed Center. [Available online at www.dtcenter.org/com-GSI/users/docs/index.php.]
- , cited 2011b: WRF for hurricanes: Documents and publications. Developmental Testbed Center. [Available online at www.dtcenter.org/HurrWRF/users/docs/index.php.]
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162.
- Goerss, J. S., 2007: Prediction of consensus tropical cyclone track forecast error. *Mon. Wea. Rev.*, **135**, 1985–1993.
- Hakim, G. J., 2003: Developing wave packets in the North Pacific storm track. *Mon. Wea. Rev.*, **131**, 2824–2837.
- Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 668–688.
- HFIP, cited 2012: HFIP annual report. Hurricane Forecast Improvement Project. [Available online at www.hfip.org/documents/reports2.php.]
- Hoskins, B. J., and T. Ambrizzi, 1993: Rossby wave propagation on a realistic longitudinally varying flow. *J. Atmos. Sci.*, **50**, 1661–1671.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971–1981.
- Langland, R., M. Shapiro, and R. Gelaro, 2002: Initial condition sensitivity and error growth in forecasts of the 25 January 2000 East Coast snowstorm. *Mon. Wea. Rev.*, **130**, 957–974.
- NCAR, 2011: List of retrospective test cases for 2011 HFIP Stream 1.5 candidate models. Research Applications Laboratory, 2 pp. [Available online at www.ral.ucar.edu/projects/hfip/includes/2011_stream_1.5_test_cases.pdf.]
- NHC, cited 2011: National Hurricane Center forecast models. [Available online at www.nhc.noaa.gov/modelsummary.shtml.]
- NOAA SAB, cited 2006: Hurricane Intensity Research Working Group majority report. NOAA Science Advisory Board, 66 pp. [Available online at www.sab.noaa.gov/Reports/HIRWG_final73.pdf.]
- , cited 2007: Response to NOAA Science Advisory Board Hurricane Intensity Research Working Group. [Available online at www.sab.noaa.gov/Reports/hirwg/2010/SAB_Nov07_HFIP_Response_to_HIRWG_FINAL.pdf.]
- NOARL, 1992: The NOGAPS forecast model: A technical description. Naval Oceanographic and Atmospheric Research Laboratory, Final Rep. AD-A247 216, DTIC 92-06202, 223 pp. [Available online at www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA247216&Location=U2&doc=GetTRDoc.pdf.]
- NRL, cited 2011: COAMPS-TC tropical cyclone prediction and verification. [Available online at www.nrlmry.navy.mil/coamps-web/web/tc?&spg=2.]
- NSF, 2007: Hurricane warning: The critical need for a national hurricane research initiative. National Science Foundation Rep. NSB-06-115, 36 pp. [Available online at www.nsf.gov/nsb/committees/archive/hurricane/initiative.pdf.]
- OFCM, 2007: Interagency strategic research plan for tropical cyclones—The way ahead. OFCM Rep. FCM-P36-2007, 270 pp. [Available online at www.ofcm.gov/p36-isrtc/fcm-p36.htm.]
- Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747–1763.
- PSU, cited 2011: PSU WRF/EnKF Real-time Atlantic hurricane forecast. [Available online at http://hfip.psu.edu/realtime/AL2011/forecast_track.html.]
- Purser, J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003a: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1524–1535.
- , —, —, and —, 2003b: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Wea. Rev.*, **131**, 1536–1548.
- Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121–150.
- Reynolds, C. A., M. S. Peng, and J.-H. Chen, 2009: Recurving tropical cyclones: Singular vector sensitivity and downstream impacts. *Mon. Wea. Rev.*, **137**, 1320–1337.

- Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Weng, Y., and F. Zhang, 2012: Assimilating airborne Doppler radar observations with an ensemble Kalman filter for cloud-resolving hurricane initialization and prediction: Katrina (2005). *Mon. Wea. Rev.*, **140**, 841–859.
- WISC, cited 2011: UWNMS model configuration for 2011 HFIP demo. [Available online at <http://cup.aos.wisc.edu/will/HFIP/config.html>.]
- Wu, W.-S., J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.
- Zhang, F., Y. Weng, J. Gamache, and F. Marks, 2011: Performance of convection-permitting hurricane initialization and prediction during 2008–2010 with ensemble data assimilation of inner-core airborne Doppler radar observations. *Geophys. Res. Lett.*, **38**, L15810, doi:10.1029/2011GL048469.