

# Impact of Assimilating Aircraft Reconnaissance Observations on Tropical Cyclone Initialization and Prediction Using Operational HWRF and GSI Ensemble–Variational Hybrid Data Assimilation

MINGJING TONG,<sup>a,b,c</sup> JASON A. SIPPEL,<sup>d</sup> VIJAY TALLAPRAGADA,<sup>a</sup> EMILY LIU,<sup>a,c</sup> CHANH KIEU,<sup>a,b</sup> IN-HYUK KWON,<sup>a,b</sup> WEIGUO WANG,<sup>a,b</sup> QINGFU LIU,<sup>a</sup> YANGRONG LING,<sup>a,b</sup> AND BANGLIN ZHANG<sup>a,b</sup>

<sup>a</sup> NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland

<sup>b</sup> I.M. Systems Group, Rockville, Maryland

<sup>c</sup> Engility at NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

<sup>d</sup> NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida

<sup>e</sup> Systems Research Group, College Park, Maryland

(Manuscript received 12 December 2017, in final form 15 May 2018)

## ABSTRACT

This study evaluates the impact of assimilating high-resolution, inner-core reconnaissance observations on tropical cyclone initialization and prediction in the 2013 version of the operational Hurricane Weather Research and Forecasting (HWRF) Model. The 2013 HWRF data assimilation system is a GSI-based hybrid ensemble–variational system that, in this study, uses the Global Data Assimilation System ensemble to estimate flow-dependent background error covariance. Assimilation of inner-core observations improves track forecasts and reduces intensity error after 18–24 h. The positive impact on the intensity forecast is mainly found in weak storms, where inner-core assimilation produces more accurate tropical cyclone structures and reduces positive intensity bias. Despite such positive benefits, there is degradation in short-term intensity forecasts that is attributable to spindown of strong storms, which has also been seen in other studies. There are several reasons for the degradation of intense storms. First, a newly discovered interaction between model biases and the HWRF vortex initialization procedure causes the first-guess wind speed aloft to be too strong in the inner core. The problem worsens for the strongest storms, leading to a poor first-guess fit to observations. Though assimilation of reconnaissance observations results in analyses that better fit the observations, it also causes a negative intensity bias at the surface. In addition, the covariance provided by the NCEP global model is inaccurate for assimilating inner-core observations, and model physics biases result in a mismatch between simulated and observed structure. The model ultimately cannot maintain the analysis structure during the forecast, leading to spindown.

## 1. Introduction

Assimilation of high-resolution data within tropical cyclones (TCs) can significantly reduce errors in TC forecasts. The potential of such data to improve TC forecasts has been known for about a decade and was first demonstrated in the literature by Zhang et al. (2009). That study used an experimental ensemble Kalman filter (EnKF) coupled with the Weather Research and Forecasting (WRF) Model to assimilate WSR-88D Doppler velocity

data from Hurricane Humberto (2007). The resultant forecast improvements were impressive, spawning a great deal of interest in the concept. Weng and Zhang (2012) followed with a study of Hurricane Katrina (2005) using the same WRF-EnKF system to assimilate airborne Doppler velocity data from the National Oceanic and Atmospheric Administration (NOAA) P-3 tail Doppler radar (TDR). The positive impact on track and intensity forecasts in that study demonstrated that inner-core, high-resolution reconnaissance could be successfully used to improve TC forecasts. Zhang et al. (2011) and Zhang and Weng (2015) showed over a number of cases with the same system that forecast errors equal to or lower than those from the National Hurricane Center (NHC) could be achieved by assimilating airborne Doppler velocity data. Weng and Zhang (2016; hereafter WZ16) followed with a very large

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Dr. Mingjing Tong, mingjing.tong@noaa.gov

DOI: 10.1175/MWR-D-17-0380.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

retrospective over a 5-yr period and found that the combination of all reconnaissance types improved forecasts of the maximum 10-m wind ( $V_{max}$ ) and center location by about 10% and 15%, respectively. Meanwhile, Aksoy et al. (2013) and Aberson et al. (2015; hereafter A15) examined 83 cases with their experimental, EnKF-based Hurricane Ensemble Data Assimilation System (HEDAS) and also found that assimilation of airborne reconnaissance observations had a significant positive impact on TC analyses and intensity forecasts. A number of other studies (e.g., Li et al. 2012; Zhu et al. 2016; Christophersen et al. 2017) have also shown the benefit of assimilating TC inner-core data over more limited samples.

Major difficulties for inner-core data assimilation (DA) in many studies include initial intensity biases and prominent spindown<sup>1</sup> of hurricanes. In A15, for example, assimilation of some reconnaissance types caused strong negative  $V_{max}$  bias that lasted for 24 h. Vukicevic et al. (2013) examined the spindown problem in great detail for HEDAS and found that forecast initialized with the most intense hurricanes exhibited the largest spindown. In another example, assimilation of reconnaissance data in WZ16 degraded  $V_{max}$  for the first 24 h. Though this has not been documented in literature, very large spindown (at times  $>25 \text{ m s}^{-1}$ ) has likewise been observed in real-time experiments with the WZ16 system. Hendricks et al. (2011) also documented substantial spindown issues for the Naval Research Laboratory's tropical cyclone version of the Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS-TC). Given that initial biases and spindown appear to be a major, common problem in TC forecasting, a significant goal of this paper is to document their prevalence and discuss potential causes as well as remedies.

This study presents results from the first and largest retrospective to assess assimilation of high-resolution, inner-core TC data into an operational model: the 2013 version (H213) of the Hurricane Weather Research and Forecasting system (HWRF) at the National Centers for Environmental Prediction (NCEP). As such, this is the first glimpse at the impact that airborne reconnaissance has had on HWRF forecasts in an operational setup and a baseline from which improvement can be expected. As we will show, these initial results presented NCEP with both benefits and challenges, and have guided the implementation of inner-core assimilation over the past few years.

---

<sup>1</sup> Spindown refers to a dramatic weakening of the simulated vortex during the first 6–12 h, which is a leading cause of negative intensity bias for strong TCs in HWRF.

While inner-core data assimilation has the potential to improve dynamical model intensity forecasts, it was a challenging issue in the HWRF implementation. Unlike the experimental results presented above, HWRF uses Gridpoint Statistical Interpolation (GSI; Wu et al. 2002), a three-dimensional variational (3DVAR) algorithm, for data assimilation. Initial experiments with numerous configurations of GSI in HWRF showed negative impacts from assimilating inner-core data. The problems stemmed from the background error covariance traditionally used in 3DVAR, which is largely isotropic and assumed to be static and constrained by large-scale geostrophic balance. With such a background error covariance, observation information cannot effectively spread in space and between physical variables in the TC vortex area during data assimilation (e.g., Fig. 8 of Lu et al. 2017a).

There are several alternatives to 3DVAR that can improve data assimilation in hurricanes. For example, the main advantage of an EnKF over the traditional variational method is that it uses short-range ensemble forecasts to estimate forecast error covariance, which is fully flow dependent and does not rely on prescribed balances that may not hold true for TCs. Another assimilation alternative is hybrid data assimilation (Lorenc 2003; Buehner 2005), which combines the advantages of the EnKF and the variational methods (Wang 2010). The hybrid method was developed and implemented in the NCEP Global Forecast System (GFS) in 2012. With the hybrid method, the flow-dependent ensemble covariance can be incorporated into the variational framework. To more effectively ingest inner-core in situ and remote sensing observations for TC applications, a GSI-based one-way hybrid data assimilation system was developed for HWRF and is used in this study.

HWRF also uses a vortex initialization (VI) procedure, which typically is not present in research assimilation systems, to help initialize the vortex and reduce forecast biases when data are sparse. The VI technique developed by Liu et al. (2012) uses storm location, intensity, and size in the Tropical Cyclone Vitals (TCVitals) reports to relocate and modify the vortex (see also Tallapragada et al. 2013). Extensive testing has shown that VI is a necessary component of HWRF, particularly when reconnaissance is not available.

The principle data of interest here are high-resolution, inner-core aircraft observations, including the NOAA P-3 TDR radial velocity data, aircraft flight-level data, and stepped frequency microwave radiometer (SFMR) wind speed retrievals at 10-m altitude. HWRF is one of very few models utilizing these new sources of data in real time.

The data are also assimilated into the WZ16 system, which is based on the WRF-ARW Model. In this study, detailed implementation and experiments with the aircraft reconnaissance observations for the H213 will be documented.

The manuscript is constructed as follows. The H213 implementation, the one-way hybrid data assimilation system, and the design of the data assimilation experiments for the reconnaissance data impact study will be described in section 2. Section 3 shows verification of the data assimilation experiments spanning the 2008–12 hurricane seasons, as well as detailed analysis of the data impacts. Concluding remarks are given in the final section.

## 2. Methodology and experimental design

### a. HWRF Model

HWRF runs separately for each storm as a triply nested model with domains that follow the storm center. The H213 domain configuration has 43 vertical levels and a 50-hPa model top. The sizes of the forecast domains are shown in Fig. 1. The horizontal grid spacing<sup>2</sup> of the outer, middle, and inner domains is, respectively, 0.18°, 0.06°, and 0.02° (roughly 27, 9, and 3 km) in the rotated latitude–longitude E grid used by the HWRF (WRF-NMM) dynamic core. Lateral boundary conditions are taken from the GFS forecasts and updated every 6 h.

The model physical parameterizations include the NCEP GFS planetary boundary layer (PBL) scheme based on observational findings (Gopalakrishnan et al. 2013; Zhang et al. 2015), improved Geophysical Fluid Dynamics Laboratory (GFDL) surface physics, improved Ferrier microphysics (Ferrier 1994), and implementation of the revised GFS shallow and deep convective parameterization (Han and Pan 2011) in the outer domain. The cloud-permitting 3-km nest is configured to explicitly resolve convection in the inner core of TCs to be consistent with the higher-spatial-resolution grids. H213 is coupled to the Princeton Ocean Model for Tropical Cyclones (Mellor 2004) in the North Atlantic Ocean (NATL) and the eastern North Pacific Ocean (EPAC).

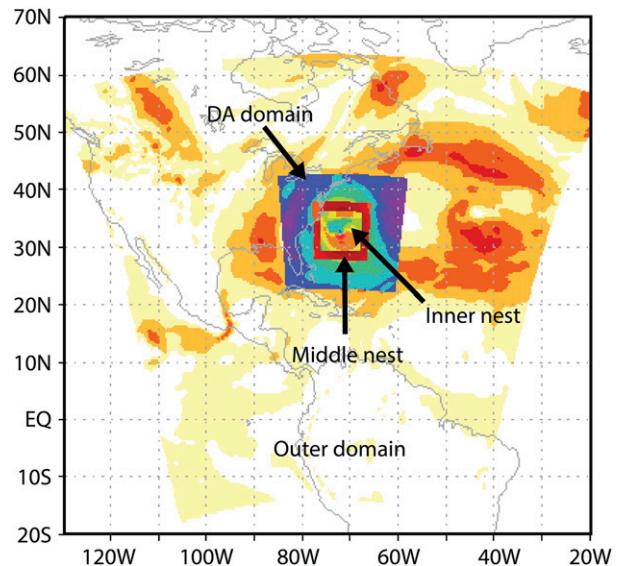


FIG. 1. HWRF forecast and DA domains.

### b. Vortex initialization

Since operational HWRF forecasts must cover the whole lifetime of a TC, which often falls outside the range of reconnaissance missions, VI is still an important component of the HWRF system. In basins with data assimilation (the North Atlantic and eastern North Pacific as of the time of this publication), VI preconditions the previous 6-h forecast to provide a first guess for GSI. In other basins, VI provides the initial conditions from which the subsequent forecast is run.

The VI consists of vortex relocation (VR) and vortex adjustments (VAs), which can be independently implemented. In general, VI removes the vortex from the previous 6-h HWRF forecast, makes adjustments to both vortex size and intensity (if VA is turned on), and implants the adjusted vortex into the NCEP Global Data Assimilation System (GDAS) 6-h forecast interpolated to a  $30^\circ \times 30^\circ$  domain used for VI (with its own vortex removed). During VA, the vortex is first stretched or compressed to more closely match the operational TCVitals assessment of the radius of maximum winds (RMW) and either the four-quadrant maximum radius of 34-kt ( $1 \text{ kt} = 0.51 \text{ m s}^{-1}$ ) winds (R34) or the radius of the outer closed isobar (ROCI). When matching the maximum R34, the model R34 (calculated using the same definition as in TCVitals) is calculated by averaging the maximum radius of 34-kt wind in each quadrant. After size adjustment, the axisymmetric component is rebalanced according to gradient wind balance, and water vapor is adjusted assuming relative humidity is unchanged during rebalancing. Next, the

<sup>2</sup>The horizontal grid spacing of the staggered E grid is the distance between adjacent mass and wind points. The conventional grid spacing is the shortest diagonal distance between two adjacent mass or wind points (Fig. 5.2 of Janjic et al. 2010). The conventional grid spacing is approximately equal to  $111.0(dx^2 + dy^2)^{1/2}$  km, where  $dx$  and  $dy$  are the horizontal grid spacing of the staggered E grid in degrees.

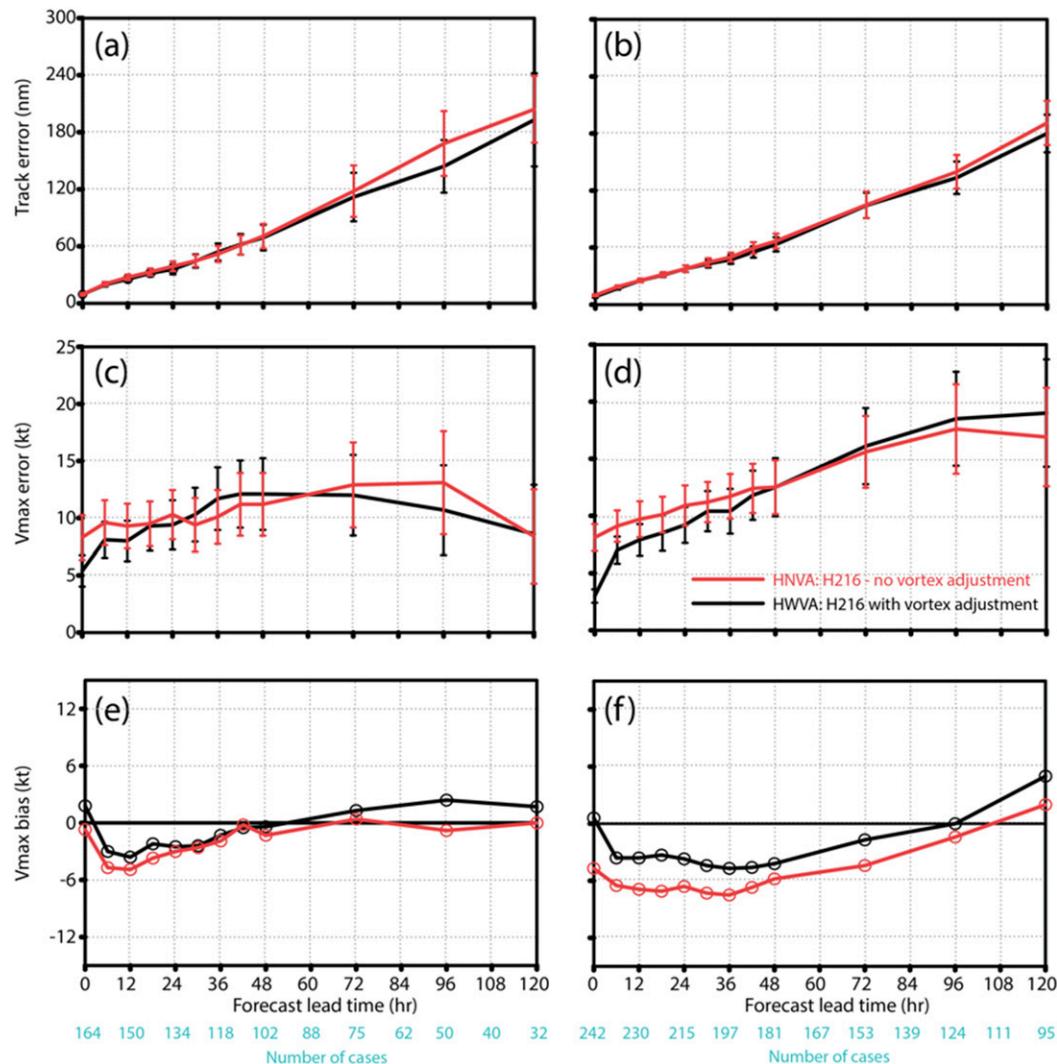


FIG. 2. Mean absolute errors for (a),(b) track (n mi; 1 n mi = 1.852 km) and (c),(d) maximum 10-m wind speed ( $V_{\max}$ , kt) as well as (e),(f) the  $V_{\max}$  bias (kt) in experiments with (HWVA) and without (HNVA) vortex adjustment for 2014 TCs in the (left) NATL and (right) EPAC basins. The error bars denote the 95% confidence interval.

wind speeds throughout the vortex are adjusted so that the maximum 10-m wind speed ( $V_{\max}$ ) more closely matches that from TCVitals. Finally, a second round of balancing and moisture adjustments is required after the wind adjustment.

Note that limits on size and intensity adjustments are applied during VA to account for uncertainties in the assessment of storm state (e.g., Landsea and Franklin 2013; Trahan and Sparling 2012) and model representativeness errors. This aspect of VA is similar to how observation errors are taken into account in data assimilation. The exact implementation of VI is described more thoroughly in Tallapragada et al. (2013).

Recent tests and publications have illuminated the contribution of VI to reducing HWRf forecast errors.

Lu et al. (2017b) showed that VR is beneficial even when high-resolution, inner-core data are available. Without relocation, analysis increments tend to be highly asymmetric and unrealistic, resulting in large forecast errors. Meanwhile, though Lu et al. (2017b) showed that VA is not strictly necessary when high-resolution data are available, that is not often the case. Figure 2 illustrates the impact of VA on track and intensity errors over the entire 2014 season in the NATL and a large part of that season in the EPAC basins. The most obvious impact from the 2014 sample is a large reduction in error of the initial  $V_{\max}$  in both basins that is associated with a reduction in negative bias. In NATL, the difference is only statistically significant at the analysis time, though on average, the benefit remains through 24 h. After 24 h, the

results are mixed such that the net impact on the Vmax forecast is weakly positive. Meanwhile, VA results in statistically significant improvements in EPAC Vmax for the first 12 h, and the positive benefit remains through 48 h. This improvement is associated with a large reduction in weak intensity bias. There is some degradation in EPAC Vmax at longer lead times due to VA, though it is not statistically significant due to large variance. Finally, VA appears to result in a small, statistically insignificant reduction of track error at most lead times in both basins. Furthermore, additional experiments with both VR and VA turned off reveal forecast error statistics very similar to the experiments without vortex adjustment (HVNA) in Figs. 2a, 2c, and 2e (not shown). Although the initial center location error nearly doubles with VR turned off, this does not appear to impact forecast error as much as VA does. Similar results to those shown here have been seen in other tests with different storms and versions of HWRF.

#### c. One-way hybrid ensemble–variational data assimilation system for HWRF

The H213 Data Assimilation System (HDAS) is an economic one-way hybrid data assimilation system based on GSI 3DVAR-based ensemble–variational hybrid data assimilation. The extended control variable method (Lorenc 2003; Buehner 2005) is used to incorporate ensemble covariance in the variational data assimilation framework [mathematical details about the implementation of the algorithm in GSI can be found in Wang (2010)]. Unlike the two-way coupled hybrid data assimilation system developed for the NCEP GFS (Hamill et al. 2011; Wang et al. 2013; Kleist and Ide 2015), the EnKF component was not included in HDAS to update the ensemble. Instead, the 80-member GDAS ensemble directly provided flow-dependent background error covariance. The horizontal resolution of the GDAS ensemble is T254, or approximately 78 km at the equator. Data assimilation is performed every 6 h on the HWRF outermost forecast domain and an inner  $20^\circ \times 20^\circ$  domain with a horizontal grid spacing of  $0.02^\circ$  (Fig. 1). The inner DA domain, which is used for vortex-scale analysis, is larger than the forecast nests so that the analysis can be interpolated to the two forecast nests as well. The DA domain is also sufficiently large to encompass a transition zone on the outer domain, which is applied to prevent discontinuities between the inner and outer domains. A procedure called “first guess at appropriate time” (FGAT) (Rabier et al. 1998; Lawless 2010) is used to more accurately calculate innovations for the observations distributed within the 6-h assimilation window.

The background error covariance in HDAS is treated differently in the inner and outer domains. In the outer-domain analysis, the covariance is a blend of 25% static and 75% ensemble covariance, whereas the weighting factor for ensemble covariance is increased to 80% in the inner domain because the static covariance is less effective in representing the balance of TC vortices. The horizontal localization scales (Gaussian *e*-folding decorrelation length scale) used for assimilation in the outer and inner analysis domains are 600 and 150 km, respectively. The vertical localization is 0.5 scale heights (i.e., natural log of the pressure) for the outer domain, but for the inner analysis domain, the vertical localization depends on storm intensity. For storms weaker than the category-1 threshold, the vertical decorrelation scale is 10 model levels, and it is 20 model levels for storms of greater intensity. The static covariance estimated by the NMC method (Parrish and Derber 1992) is latitude and height dependent. Both the covariance blending ratios and the localization length scales were chosen based on forecast impact tests.

Two outer loops with 50 inner iterations each are used for GSI minimization. For nonlinear observation operators, the tangent-linear approximation is generally used, and the Jacobians are updated at each outer loop. For simple nonlinear observation operators that are not expensive to update with each inner iteration, such as the observation operator for SFMR-derived surface wind speeds,<sup>3</sup> the option of using a nonlinear operator is also included in GSI. In this study, the nonlinear operator for SFMR surface wind speeds is used.

The analysis variables include streamfunction, unbalanced velocity potential, unbalanced temperature, unbalanced surface pressure, normalized relative humidity, and the extended control variable  $\alpha$ . Model state variables updated include zonal and meridional wind speed, temperature, specific humidity, hydrostatic pressure depth, and nonhydrostatic pressure.

#### d. Observations and experiment design

This study focuses on impacts of two types of high-resolution, inner-core observations: TDR radial velocity observations from the NOAA P-3 aircraft and high-density observations (HDOBs), which include flight-level wind, temperature, moisture measurements, and SFMR-derived surface wind speeds from NOAA P-3

<sup>3</sup> The observation operator for SFMR surface wind speed data is expressed as  $(u^2 + v^2)^{1/2}$ , where  $u$  and  $v$  are model zonal and meridional wind at the observation location multiplied by the 10-m wind reduction factor. The 10-m wind reduction factor is the ratio of the model output wind speed at 10-m height to the model first-level wind speed.

and Air Force Reserve Command (AFRC) C-130 aircraft. The performance of HWRF with assimilation of high-resolution, inner-core observations is evaluated based on four experiments (Table 1) conducted for applicable storms spanning five hurricane seasons from 2008 to 2012. The control experiment (HCTL) does not assimilate HDOB or TDR data, while the remaining experiments assimilate TDR radial velocity (HTDR), HDOB data (HDOB), and both TDR and HDOB (HARC). The storms and the number of cycles with TDR data and HDOB available are listed in Table 2.

All experiments assimilate a large variety of conventional data contained in standard NCEP Prepared Binary Universal Form for the Representation of meteorological data (PREPBUFR)<sup>4</sup> files. GPS dropwindsonde (dropsonde hereafter) wind and thermodynamic profiles, which are assimilated into other NCEP operational models, undergo a quality control (QC) procedure to remove certain data from the tropical cyclone vortex. In particular, dropsonde surface pressure data within 300 km of the storm center and dropsonde wind data within a radius of 111 km or 3 times the RMW, whichever is larger, are excluded from assimilation. Experimentation has shown a generally negative impact of those data due to both dropsonde drift<sup>5</sup> and to issues discussed in the next section with assimilating inner-core data in the HWRF system for strong storms. In addition to dropsonde data, other conventional observations assimilated are radiosonde, aircraft reports, surface ship and buoy observations, surface observations over land, pilot balloon winds, wind profilers, weather surveillance Doppler radar velocity azimuth display winds, scatterometer winds, and GPS-derived integrated precipitable water.

For HCTL, data assimilation is performed only on the HWRF outer domain with conventional observations. Data assimilation in HCTL mainly aims to improve initialization of the large-scale TC environment, and the TC vortex is initialized through the VI procedure. In the inner-core assimilation experiments, when high-resolution, inner-core observations are available, data

TABLE 1. List of inner-core DA experiments.

Experiment	Configuration and observations assimilated
HCTL	Conventional data on outer domain followed by vortex initialization (no inner-core DA)
HTDR	Add TDR radial velocity ( $V_r$ ) data
HDOB	Add HDOB data HDOB data include flight-level wind ( $u, v$ converted from wind speed and direction), temperature, specific humidity (converted from dewpoint), and SFMR retrieved surface wind speed
HARC	Add both TDR $V_r$ and HDOB data

assimilation is performed on both the outer domain and the inner analysis domain, with the inner-core observations assimilated only in the inner domain. To be consistent with the data assimilation on the outer domain, conventional data are also assimilated in the inner analysis domain when inner-core observations are available.

NCEP started to receive real-time TDR data beginning in the 2010 hurricane season. The TDR data were collected by NOAA's Aircraft Operations Center (AOC) and preprocessed by the Hurricane Research Division (HRD) of NOAA's Atlantic Oceanographic and Meteorological Laboratory (AOML). Major quality control (Gamache 2005) of the TDR radial velocity data conducted on board the P3 aircraft includes 1) removing the projection of the aircraft motion on the observed Doppler velocity; 2) removing the reflection of the main lobe and side lobes off the sea surface; 3) removing noise; and 4) velocity unfolding. The radial velocity data were averaged over eight gates along the radial direction and processed into BUFR format, and data within the 6-h window centered on the analysis time were dumped into one BUFR file. The storms with BUFR format of TDR data available are Earl (2010), Irene (2011), Rina (2011), Isaac (2012), Leslie (2012), and Sandy (2012). For all other storms with TDR data, we used superobs (Weng and Zhang 2012; Aksoy et al. 2013) generated by HRD. The superobs were processed into 6-h files with the same cutoff time as the BUFR dumps. The same data thinning and quality control were conducted for both BUFR data and superobs. The observation error, which includes representativeness error, is set to be  $5 \text{ m s}^{-1}$  and is gradually inflated to  $10 \text{ m s}^{-1}$  when the difference between the observation and background (i.e., innovation) is greater than  $10 \text{ m s}^{-1}$ . Experiments were conducted to compare the assimilation of TDR BUFR data and TDR superobs, and the results were found to be comparable.

The HDOB data were also processed into 6-h data files in the BUFR format. The flight-level temperature, dewpoint temperature, and wind direction–speed were

<sup>4</sup> PREPBUFR files used in HWRF assimilation contain the same BUFR of meteorological data format of conventional data that is used in GDAS. This format contains conventional observations in 6-h time windows centered at each synoptic time.

<sup>5</sup> Dropsonde data are transmitted in the TEMPDROP format, the body of which contains only the release point (Aberson et al. 2017). Thus, as of the time of this publication, NCEP practice is to assimilate all data from a single dropsonde at the same horizontal location regardless of drift (i.e., advection) of the dropsonde as it falls.

TABLE 2. The 2008–12 storms (storm number and name) considered for the experiments and number of cycles with HDOB and TDR data available of each storm.

Year	Storm No.	Name	HDOB	TDR	Year	Storm No.	Name	HDOB	TDR
2008	2	Bertha	7	0	2011	1	Arlene	5	0
	3	Cristobal	9	0		2	Bret	9	0
	4	Dolly	15	6		4	Don	9	0
	5	Edouard	9	0		5	Emily	12	0
	6	Fay	23	3		7	Gert	4	0
	7	Gustav	28	5		8	Harvey	4	0
	8	Hanna	22	0		9	Irene	29	7
	9	Ike	30	7		13	Lee	4	2
	11	Kyle	10	4		14	Maria	5	0
	15	Omar	7	0		15	Nate	12	0
	17	Paloma	13	3		16	Ophelia	14	1
Total	11		172	28	18	Rina	15	3	
2009	3	Bill	19	4	Total	11		122	13
	5	Danny	11	5	2012	2	Beryl	8	0
	6	Erika	9	0		4	Debby	13	0
	11	Ida	14	0	5	Ernesto	25	0	
Total	4		53	9	9	Isaac	31	9	
2010	7	Earl	26	12	12	Leslie	3	3	
	8	Fiona	12	0	17	Rafael	11	2	
	11	Igor	11	0	18	Sandy	26	8	
	13	Karl	11	0	Total	7		117	22
	15	Matthew	4	0					
Total	5		64	12					
Total number of storms: 39			HDOB: 528			TDR: 84			

converted to virtual temperature, specific humidity, and zonal and meridional wind speed, respectively, in GSI before they were assimilated. The observation errors assigned for dropwindsonde data were used for flight-level data. Like all other types of conventional observations, the observation errors for dropwindsonde data are retrieved from an observation error table, where the observation errors are defined on pressure levels from 1100 to 0 hPa with 50-hPa interval. The observation errors for flight-level temperature and wind observations between 800 and 600 hPa change from roughly  $1.2^\circ$  to  $0.6^\circ$  and  $4.4$  to  $3.7 \text{ m s}^{-1}$ , respectively. For flight-level moisture observations, the error is 20% in terms of relative humidity. For SFMR wind speeds, the quality control marks transmitted with the data are used to determine if a datum should be used or tossed (Uhlhorn et al. 2007). The observation error for SFMR wind speed is set to be  $5 \text{ m s}^{-1}$ . All observation errors are subject to adjustment through quality control.

### 3. Results

#### a. Impact of TDR data assimilation

The impact of assimilating TDR radial velocity data depends on the forecast lead time and metrics examined. The left column of Fig. 3 shows a homogeneous comparison

of HCTL and HTDR (along with all other experiments) for the cases listed in Table 2, beginning with the cycle for which the TDR data are first available and continuing through the end of the storm. In Fig. 3a, HTDR possesses similar track forecast errors to the control experiment at all forecast lead times. Meanwhile, assimilating TDR data improves the intensity forecast errors at statistically insignificant levels after the first 18 h (Fig. 3c). The absolute intensity error reduction with the TDR assimilation appears to be a result of reduced bias since HCTL tends to have more noticeable positive intensity bias (Fig. 3e; black curve) than that seen in the HTDR experiments (Fig. 3e; red curve).

Stratifying error characteristics from Fig. 3 by storm intensity (Fig. 4) reveals that the TDR assimilation is more beneficial to initially weaker storms, but it appears to be detrimental to major hurricanes. For weak storms (i.e., storms of tropical storm strength and weaker), assimilating the TDR data improves both track and intensity forecasts. In particular, data assimilation is able to correct the large positive intensity bias that occurs when VI is used to initialize weak storms in HCTL (e.g., Fig. 4g). Though none of the improvements here are statistically significant, the improvement at 24 h does approach the 95% confidence level. For category 1 and 2 hurricanes, there is substantially less benefit from TDR data assimilation. The HTDR intensity forecast errors

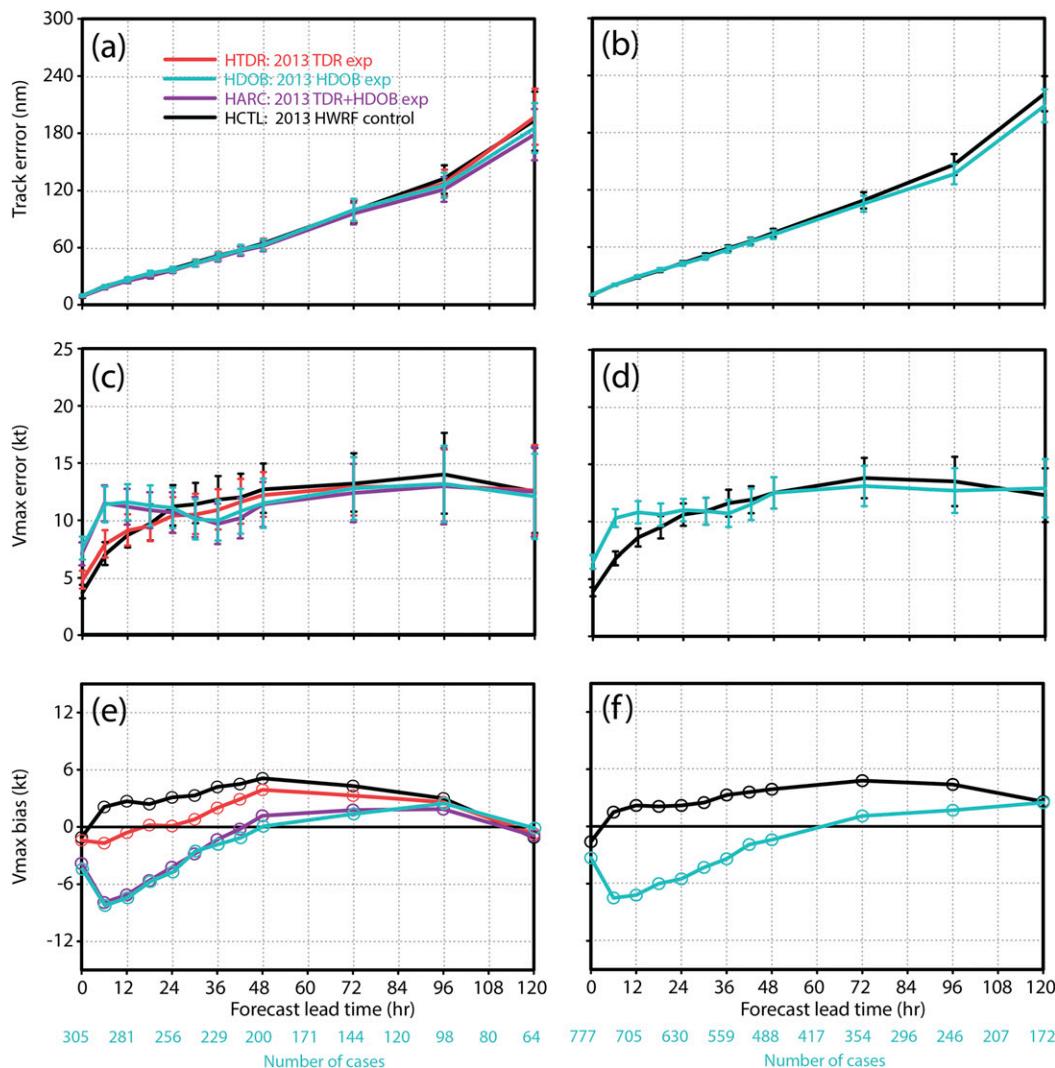


FIG. 3. Mean absolute errors of (a),(b) track (n mi) and (c),(d)  $V_{max}$  (kt) as well as (e),(f) the  $V_{max}$  bias (kt). (left) A homogeneous comparison of the four experiments in Table 1 for TCs with TDR data available; (right) a comparison between HCTL and HDOB for TCs with HDOB data available.

are commensurate with those in HCTL, and the track errors are worse, though the sample size at longer lead times is very small. Finally, major hurricane cases in HTDR suffer from a prominent negative bias through the short-term intensity forecast. On the other hand, HCTL has lower  $V_{max}$  error for major hurricane cases than for weaker systems.

To further demonstrate the impact of the TDR data on individual cases, Fig. 5 shows single-cycle track and intensity forecasts from HTDR and HCTL for a storm in each group list in Fig. 4. Note that HWRf has been cycled for at least 5 days for the cases shown in Fig. 5. For Tropical Storm Isaac, the track forecast in the TDR data assimilation experiment shows slightly smaller cross-track errors than that in the control

experiment. For Hurricanes Ike and Earl, track is commensurate in HCTL and HTDR within the first 3 days. As can be seen from Figs. 5d and 5e, the control experiment has a positive intensity bias of more than 20 kt for Isaac and Ike, but this error is largely corrected when the TDR data are assimilated. For Earl, neither the control experiment nor the TDR data assimilation experiment captures the intensification of the storm. The predicted intensity in HCTL is weaker than the intensity ( $V_{max}$ ) of the best track, and the forecast is even worse in HTDR due to severe spin-down in the first 6 h that causes a large negative  $V_{max}$  bias.

The tendency for HTDR intensity forecasts to be weaker than those in HCTL is largely a result of analysis

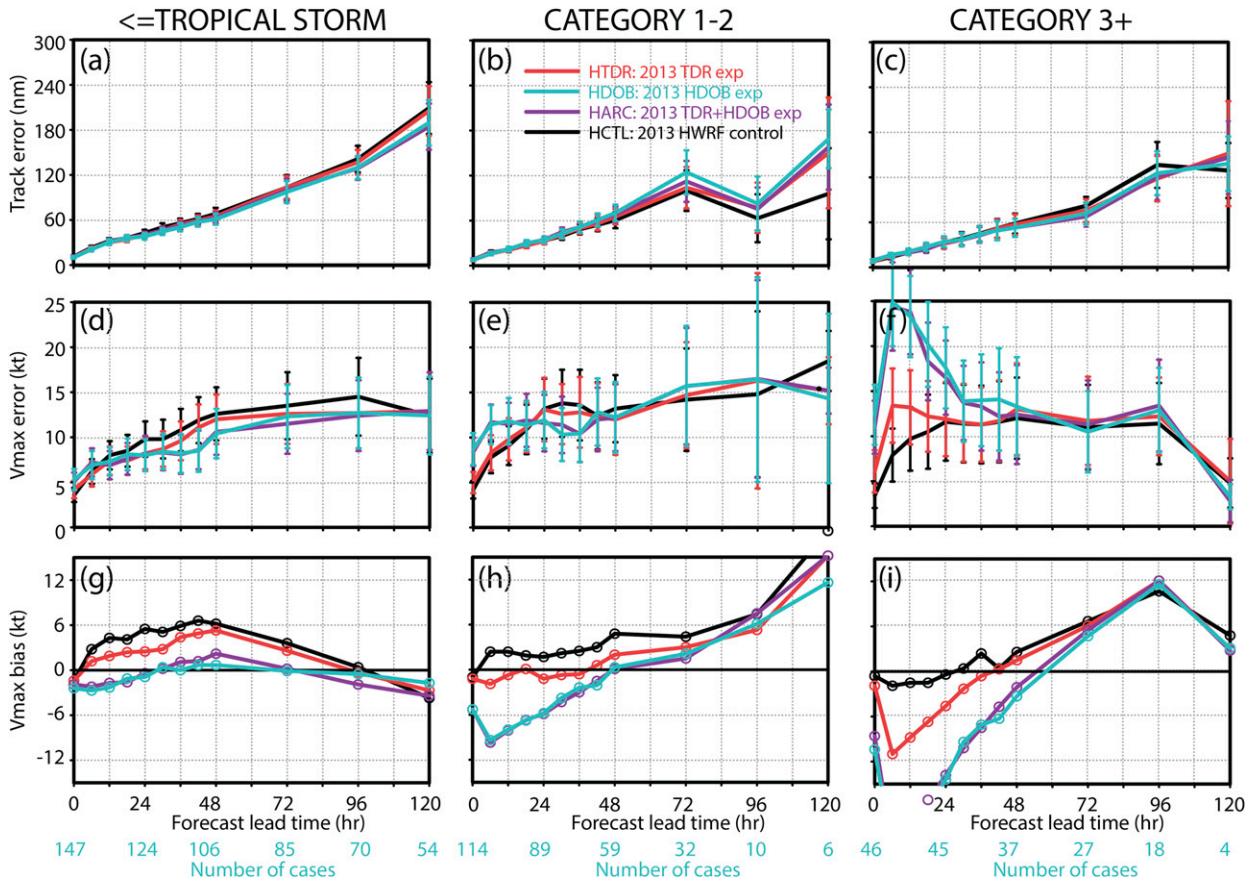


FIG. 4. As in the left column of Fig. 3, but the statistics are shown for (left) tropical storms, (middle) category 1–2 hurricanes, and (right) major hurricanes.

increments that weaken the first guess provided by VI. This is demonstrated in Fig. 6, which shows zonal wind, temperature, and moisture analysis increments in center-crossing south–north vertical cross sections for the storms from Fig. 5. For example, Fig. 6a shows a positive zonal wind increment near and north of the center of Isaac that is largest around 400 hPa. The increment generally opposes the first-guess winds, thus weakening the circulation, and also results in a northward-tilting analysis vortex. Assimilating the TDR data also reduces the strength of Ike and Earl (Figs. 6b,c), though the impact on the Earl case is much more dramatic. Note that the temperature and specific humidity increments near the storm center are introduced partly through assimilating dropsonde observations, which are available below 600 hPa, and partly through the cross-covariance and the assimilation of TDR data. Assimilation of the inner-core observations cools the upper-level center of Isaac but warms the middle levels. For Ike and Earl, the temperature increments are negative near the storm center at all levels below 200 hPa, and the

corresponding specific humidity increments around the storm center are similarly negative for all three cases. The tendency of analysis increments to cool and dry the inner core should tend to reduce the tendency to intensify in the HTDR experiment.

Figure 7 further illustrates the impact of TDR assimilation by comparing the azimuthally averaged temperature and wind structures at initial time from HTDR and HCTL. In terms of the primary circulation, all three storms in HTDR (Figs. 7d–f) are generally weaker and shallower than they are in HCTL (Figs. 7a–c). The azimuthal mean radial inflow and outflow of the Isaac and Earl cases are also significantly weaker in HTDR than they are in HCTL (not shown). Meanwhile, in HTDR, the warm core of Isaac is at a lower altitude, and that of Ike is weaker. Similar conclusions can also be obtained regarding the moisture fields (not shown). Ultimately, it appears that the weaker and shallower TC structures in HTDR keep storms in that experiment weaker than in HCTL.

A major reason for the consistently weaker analyses in HTDR is that the first-guess vortex is often stronger

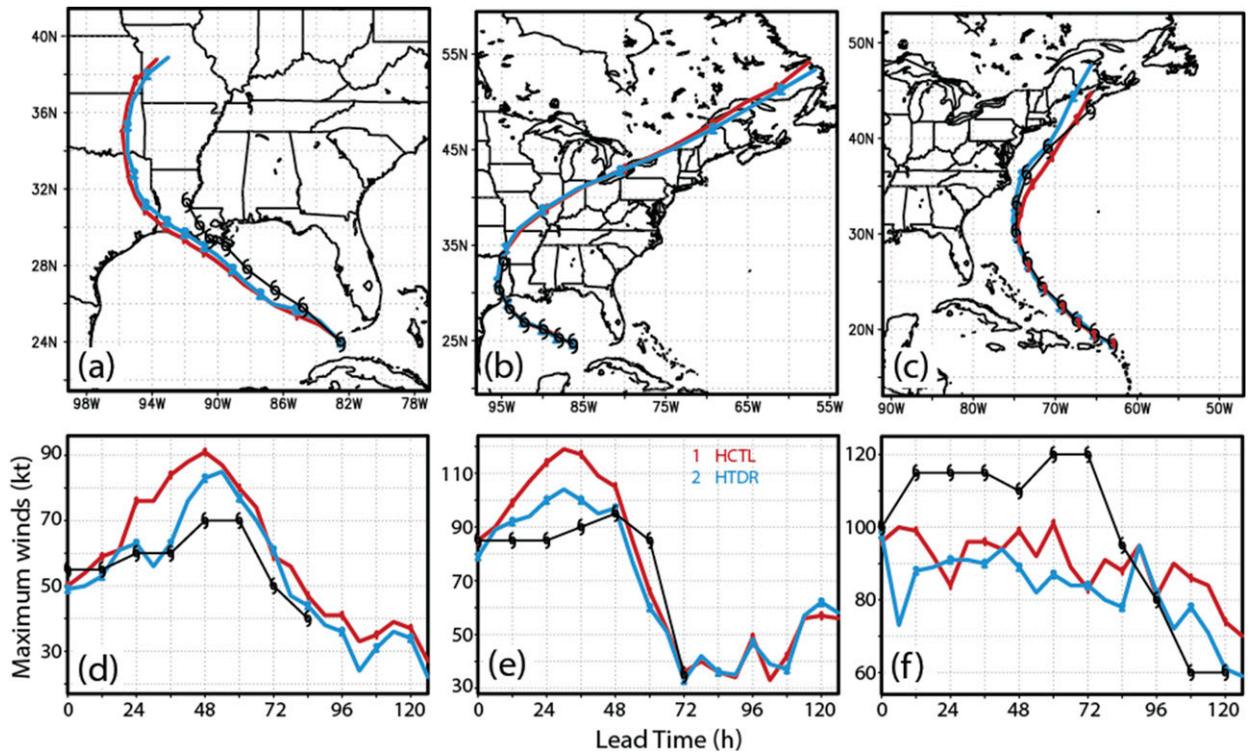


FIG. 5. Forecasts of (a)–(c) track and (d)–(f) intensity (kt) in HCTL (red) and HTDR (blue) from: (a),(d) 0000 UTC 27 Aug 2012 for Isaac; (b),(e) 0000 UTC 11 Sep 2008 for Ike; and (c),(f) 1200 UTC 30 Aug 2010 for Earl; in comparison to the NHC best track estimate (black).

than what is observed by aircraft instruments. The  $V_{max}$  of the first guess may match that of the TC Vitals due to VI, but that single value cannot represent the three-dimensional wind structure and strength observed by the tail Doppler radar. For example, P3 TDR radial velocity observations and corresponding innovations are shown in Fig. 8 to illustrate this issue for major Hurricane Earl. These P3 data shown between 700 and 800 hPa were collected from 1040 to 1409 UTC 30 August 2010. The flight track of the P3 aircraft is between the positive and negative velocity data points in Fig. 8a, and the radial velocity pattern along the flight legs penetrating the storm exhibits the typical TC cyclonic circulation. However, observation innovations in Fig. 8b show an opposing pattern, which means that the observed circulation is weaker than that of the first guess. Figure 9 further captures this problem by comparing the HTDR first guess (Fig. 9a) and analysis (Fig. 9b) wind field at 1.5-km altitude with the AOML/HRD radar wind analysis (Fig. 9c; [http://www.aoml.noaa.gov/hrd/Storm\\_pages/earl2010/radar.html](http://www.aoml.noaa.gov/hrd/Storm_pages/earl2010/radar.html); Gamache 1997; Lee et al. 2003; Lorsolo et al. 2013). The first guess has much stronger winds throughout the vortex than those in the HRD wind analysis, and it also has a

larger RMW ( $\sim 60$  km) than that of the real storm ( $\sim 46$  km). The HTDR analysis vortex is clearly closer to HRD's radar wind analysis than is the first-guess vortex.

Histograms of TDR radial wind innovations of the first guess (O-F) and analysis (O-A) in Fig. 10 reveal significant biases for stronger systems that are consistent with the above analysis. These histograms show innovations for all HTDR cases, grouped by storm initial intensity, in order to give a more systematic assessment of performance. The root-mean-square (rms) of the innovations of the first guess (O-F) in Fig. 10 and later in Fig. 13 is defined as

$$\sqrt{\frac{1}{n} \sum_{1}^{n} (O - F)^2},$$

where  $n$  is the number of data samples. The same equation applies to O-A for the analysis. As seen in Figs. 10a–c, the O-F of all three categories has larger spread than O-A so that the rms of O-A is consistently smaller than that of O-F. This is expected and confirms that data assimilation draws the model wind field closer to the observations. Of particular interest is that the rms

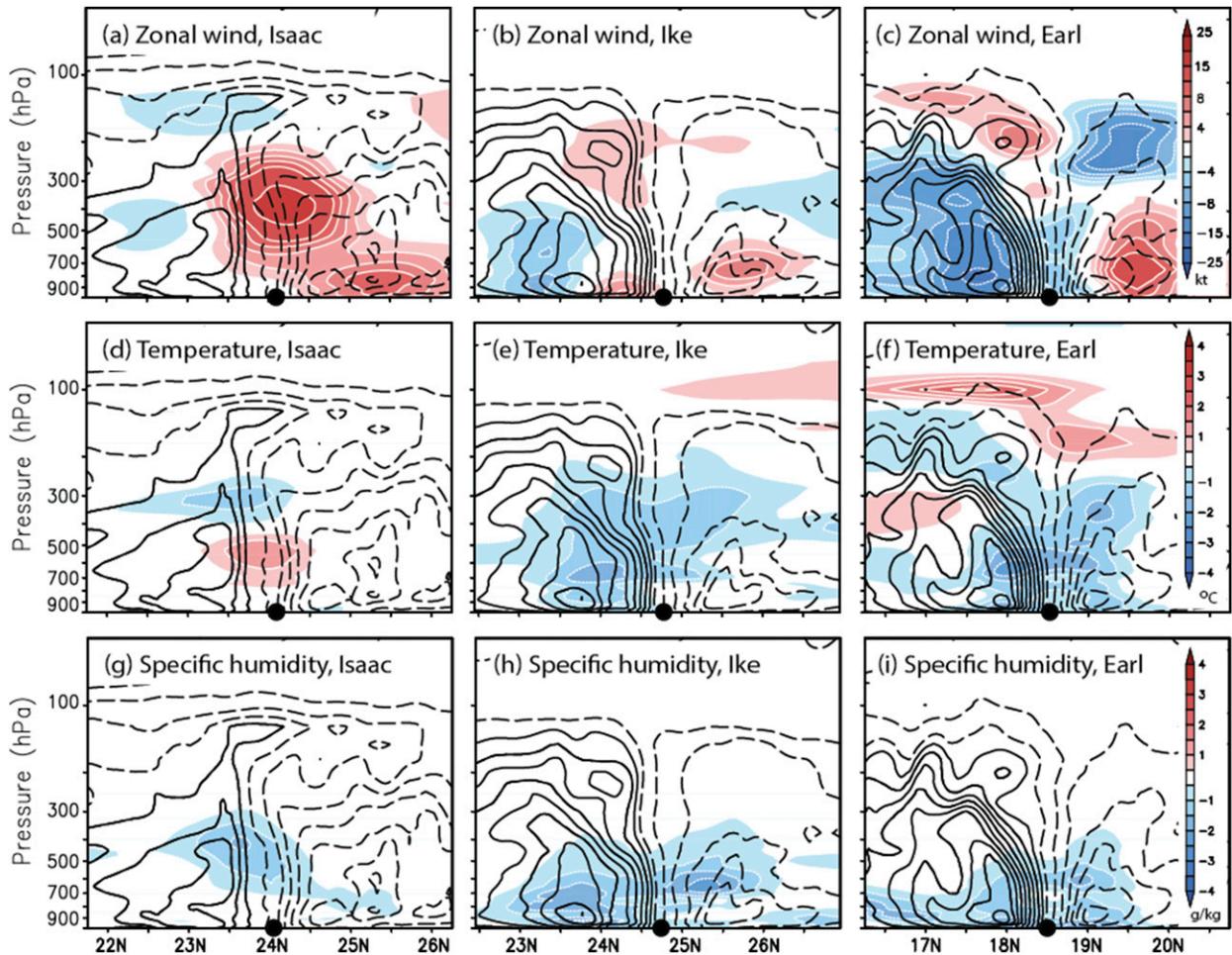


FIG. 6. South–north cross sections through the TC center showing the first-guess (contours) and analysis increments (color shades) for (a)–(c) zonal wind ( $\text{m s}^{-1}$ ), (d)–(f) temperature (K), and (g)–(i) specific humidity ( $\text{g kg}^{-1}$ ). Data are shown for Isaac at (left) 0000 UTC 27 Aug 2012, (middle) Ike at 0000 UTC 11 Sep 2008, and (right) Earl at 1200 UTC 30 Aug 2010. The TC centers are indicated by the black circles.

of O-F is larger for stronger storms, revealing that the first guess for those systems is particularly bad. The large rms error for major hurricanes is associated with a flat O-F histogram (Fig. 10c), indicating a disproportionate number of large discrepancies. Further analysis of O-F against observations in 2D histograms (Figs. 10d–f) reveals a negative correlation, consistent with the pattern in Fig. 8 that is strongest for major hurricanes. In other words, as storm intensity increases, the first-guess vortices tend to be systematically stronger than the observed vortices.

### b. Impact of HDOB data assimilation

As mentioned in the previous section, the HDOB experiment assimilates HDOB observations, which are available more often than the TDR data (Table 2). The right-hand panel of Fig. 3 shows that assimilation of the

HDOB data can help improve day-4 and day-5 track forecasts and intensity forecast after 24 h. However, the HDOB intensity forecasts are degraded during the first 24 h. In contrast to the 6-h forecast spinup and positive intensity forecast bias in the control experiment, HDOB shows 6-h spindown and large negative intensity bias before the 48-h forecast lead time (Figs. 3e,f).

We now turn to a homogeneous comparison of error statistics between HTDR and HDOB to better understand the relative impacts of TDR and HDOB data. A problem that arises when trying to compare the experiments in Figs. 3 and 4 is that HDOB data are available for many more cycles than are TDR data. Thus, to compare the impact of temporal availability with the impact of data itself, we introduce Figs. 11 and 12 to show a homogeneous sample of error characteristics for only the cycles when both HDOB and TDR data are available.

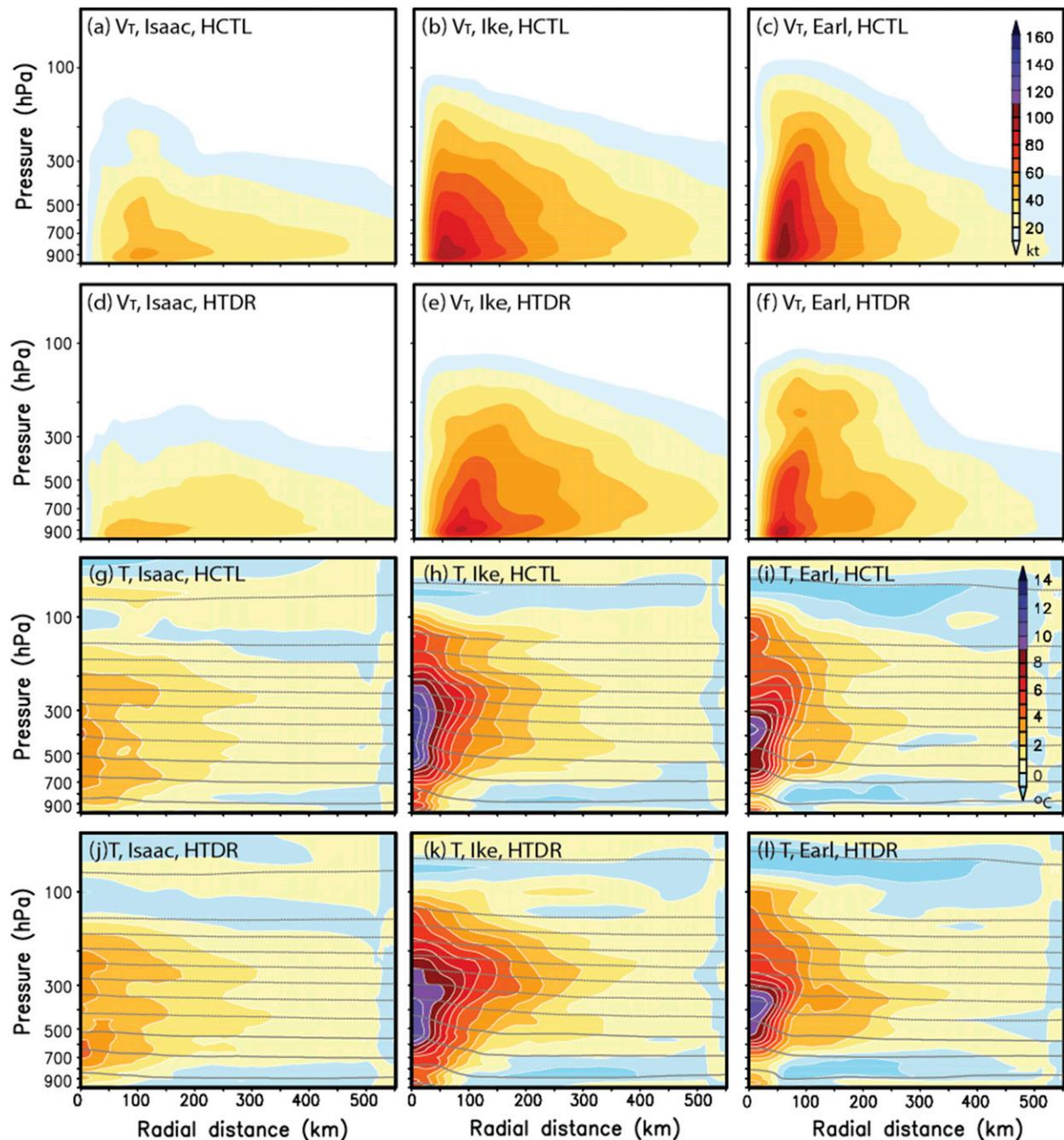


FIG. 7. A comparison of azimuthally averaged radius-height profiles in (a)–(c) and (g)–(i) HCTL, and (d)–(f) and (j)–(l) HTDR. Fields examined are the (a)–(f) tangential wind and (g)–(l) temperature (contours) and temperature anomaly (shaded). The cases shown are (left) Isaac at 0000 UTC 27 Aug 2012, (middle) Ike at 0000 UTC 11 Sep 2008, and (right) Earl at 1200 UTC 30 Aug 2010.

Some differences in error characteristics between HDOB and HTDR are a result of better HDOB data availability, while other differences can be attributed to both the data and their availability. For example, in addition to the issue discussed above, the characteristics of TDR and HDOB data are very different. TDR data

cover much of the model depth and the breadth of the vortex, but they only directly provide kinematic information. HDOB data provide both kinematic and thermodynamic information, but they are mostly at a single level (plus wind speed at the surface). In Fig. 3a, HDOB has somewhat lower track errors than HTDR at

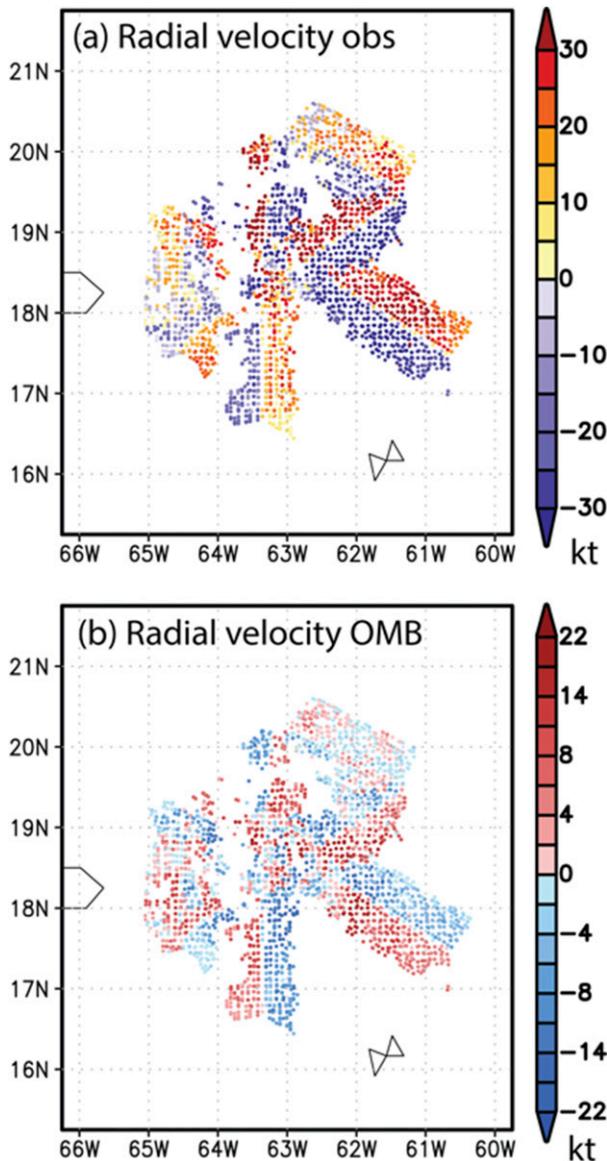


FIG. 8. (a) The TDR radial velocity observations between 800 and 700 hPa that are assimilated into the 1200 UTC 30 Aug 2010 cycle of Earl and (b) the corresponding innovations from HTDR.

120 h. However, that difference disappears if only the cycles with both data available are considered (Fig. 11a). On the other hand, the strong initial negative bias suggestive of spindown is worse in HDOB than HTDR for both the full and limited samples (cf. Figs. 3e and 11c). That being said, the difference in negative bias is greater in Fig. 3e, where the addition of many cycles in HTDR without TDR data dampens the impact in the error statistics.

When the results are binned by initial intensity, two common themes emerge. First, HDOB produces somewhat better intensity forecasts than HTDR for tropical

storms, though the only statistically significant difference is at 42 h for the full sample (Figs. 4d, 12d). In addition, the more pronounced negative intensity bias early in the HDOB forecasts comes mainly from the hurricane cycles (Figs. 4g–i, 12g–i). This suggests that the HWRF DA system has a particularly difficult time handling HDOB data in hurricanes, perhaps because they are less volumetrically distributed than TDR data.

Finally, to evaluate the potential value added by assimilating both TDR and HDOB data, Figs. 3 and 4 and Figs. 11 and 12 also overlay error statistics from HARC onto those of HCTL, HDOB, and HTDR. In general, HARC forecast errors exhibit the undesirable negative intensity bias of HDOB at early lead times, but they provide the best forecasts after 24 h (Fig. 3). In particular, the HARC intensity forecasts are better than HCTL with 95% confidence at 36 h (Figs. 3c, 11b), which represents the only statistically significant intensity improvement for the full sample. When results are binned by initial intensity (Figs. 4, 12), HARC provides the most benefit for tropical storm cycles when both datasets are available. In this circumstance, the intensity forecasts are significantly better than HCTL from 30 to 48 h.

Similar to the analysis in Fig. 10, the innovations of the HDOB data from HDOB (Fig. 13) reveal increasing biases for stronger storms. According to the rms of O-F, the first-guess fits to all HDOBs, except for flight-level moisture, are better for weaker storms than stronger storms. Consistent with the TDR histograms, histograms of SFMR wind speeds reveal a tendency for the first-guess vortex to be significantly too strong for major hurricanes (Fig. 13i). This tendency decreases for weaker storms, though it is still modestly apparent for tropical storms (Figs. 13g,h). Likewise, the flight-level zonal wind innovations tend to be much larger for major hurricanes than for weaker systems (Figs. 13j–l). The O-F distribution of flight-level temperature and specific humidity are also negatively skewed, indicating the first guess generally has positive bias in temperature and moisture fields. Note that the wind speed innovations are skewed much farther left, toward a more intense vortex, than are the temperature innovations. This suggests that not only is the first-guess vortex too strong, but it is also too strong in a dynamically inconsistent manner. This kind of bias could contribute to analysis imbalances, as discussed below.

### c. Issues associated with inner-core data assimilation for strong storms

The reconnaissance data impact analyses in previous sections reveal that the background vortices tend to be stronger and deeper than those indicated by the inner-core observations. The large difference between the

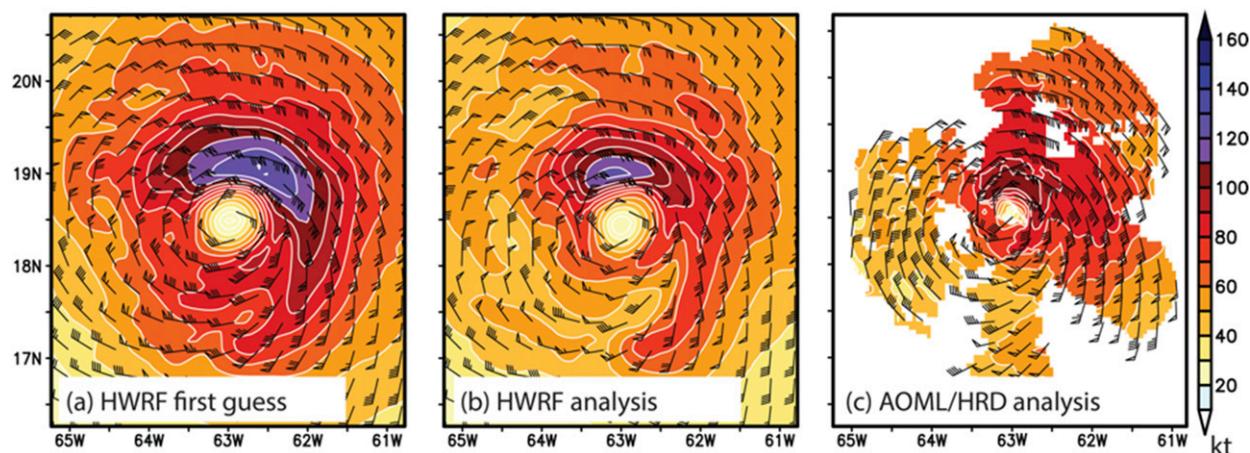


FIG. 9. The (a) first guess and (b) analysis of 1.5-km winds (kt) in Earl at 1200 UTC 30 Aug 2010 from HTDR compared with (c) AOML–HRD’s Doppler radar wind composite at 1.5-km altitude.

background vortex and the observations shown in Fig. 9 is consistently seen in other major hurricanes, such as Gustav (2008), Paloma (2008), Bill (2009), Irene (2011), and Rina (2011). In this section, we further examine the strong wind bias in the inner core of the background vortex, where it originates, and how it affects inner-core data assimilation experiments. This analysis is carried out for Hurricane Bill, another major hurricane in the 2009 season. Both the TDR and HDOB observations for Bill were available at 0000 UTC 19 August 2009, at which time  $V_{\max}$  and  $P_{\min}$  were, respectively, 105 kt and 955 hPa.

To illustrate differences between the observed structure and that from HWRf, Figs. 14a and 14b compare a vertical cross section through the TDR composite with a cross section through the 6-h HWRf forecast from the 1800 UTC 18 August cycle. There are a number of problems with the vertical structure in the forecast of Bill. First, compared to the observations, the model forecast vortex has a stronger, larger inner core, which is consistent with the previous experiments. In particular, the forecast inner-core wind speeds between 0.5 and 8 km are all too strong. Another problem is that the forecast wind speed maximum is located around 2-km altitude, which is much higher than that indicated by the TDR composite or a typical dropsonde-derived eyewall wind profile (e.g., Franklin et al. 2003). Although the vortex is too strong almost everywhere aloft,  $V_{\max}$  is 10 kt lower than that in the best track data.

The strong and large inner-core structures in the H213 forecast are compounded by the VI procedure. For example, Fig. 14c shows the vertical structure after VI (in HCTL, this is the initial condition for the next forecast cycle, and in the inner-core data assimilation experiments, it is the first guess). To be able to match the  $V_{\max}$

of the best track, VI applies an intensity correction factor everywhere in the storm. While this does bring  $V_{\max}$  closer to the best track, it also strengthens wind speeds above the surface that were already too strong to begin with. This point is emphasized in Fig. 14d, which shows that the wind structure is maintained in a 6-h forecast initialized from the vortex in Fig. 14c.

Though VI in H213 makes the vortex too strong by attempting to match  $V_{\max}$ , it appears a major contributor to this issue is an inherent model bias. This point is demonstrated by two salient properties of HCTL forecasts. First, HCTL forecasts of intense hurricanes do not significantly spin down on average, indicating that VI yields a wind structure consistent with model physics (even if it is not consistent with observations). Second, the wind profile shown in Fig. 14b remains through longer forecast lead times and is present in other storms in H213 (not shown), thus showing that it is a stable structure in the model not specifically related to initialization or a particular storm.

The above example without assimilation helps to explain why  $V_{\max}$  in H213 is too weak when inner-core data are assimilated. The problems are further illustrated with the analysis from HARC (with all recon observations assimilated) at 0000 UTC 19 August 2009, which is shown in Fig. 15. The HARC analysis in Fig. 15a matches the TDR analysis (Fig. 14a) better than the first guess does (Fig. 14c) by significantly reducing the wind speed above 2 km. Although the maximum wind speed of the analysis at 1-km altitude still matches the TDR composite, the maximum wind speeds below 1 km are too low. For example,  $V_{\max}$  of the analysis is 93 kt, which is more than 10 kt lower than that in the best track.

In addition to a biased vertical wind structure, model error that affects storm size also appears to impact inner

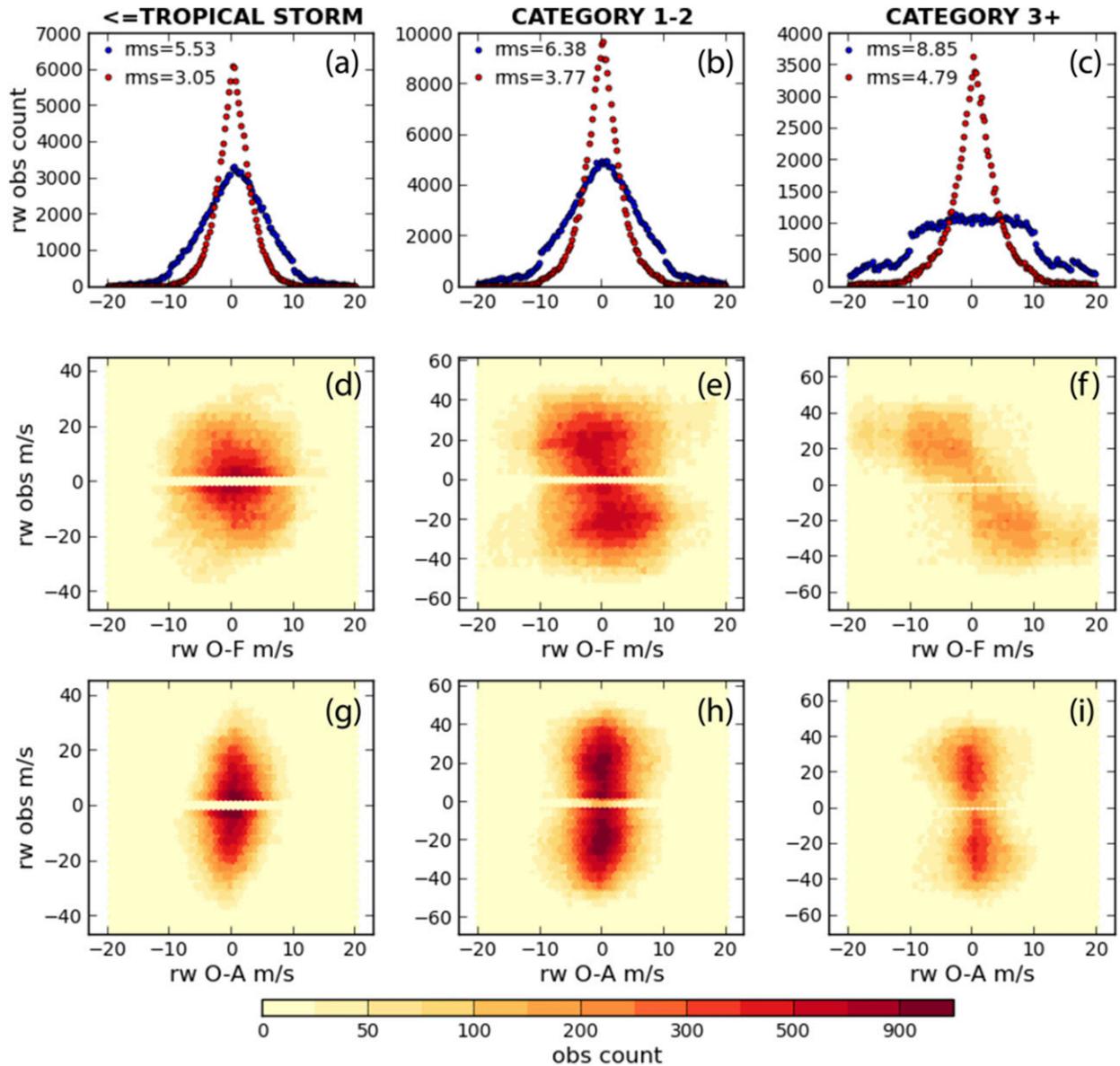


FIG. 10. Histograms of TDR radial wind (rw) innovations. (a)–(c) First guess (O-F; blue) and analysis (O-A; red) grouped by storm initial intensity as shown in Fig. 4. (d)–(f) and (g)–(i) Two-dimensional histogram of O-F and O-A compared against radial wind observations, respectively. The data are from all the cases of the HTDR experiment.

core data assimilation in H213. This is illustrated in Fig. 16, which shows wind radii errors in all experiments (averaged over four quadrants). First, HCTL has a substantial positive size bias for 50- and 64-kt wind radii. Regardless of the type of the aircraft inner-core observations assimilated, the initial 64-kt radii biases are substantially reduced with assimilation (Fig. 16f). Yet, the small initial wind radii errors in the assimilation experiments tend to increase with forecast time, which suggests that H213 intrinsically produces inner cores that are too large. Thus, the H213 configuration and

physics may not be able to support the small inner-core structures seen in observations.

Model errors that impact both storm size and vertical structure contribute to the prominent spindown observed in H213 major hurricanes. Turning back to Fig. 15, we can see that the storm weakens and the inner core expands in the short-term forecast subsequent to the 0000 UTC HARC analysis (Fig. 15b). This is consistent with the aforementioned model biases and suggests the structure imparted by observations is not consistent with model physics. Likewise, the forecast

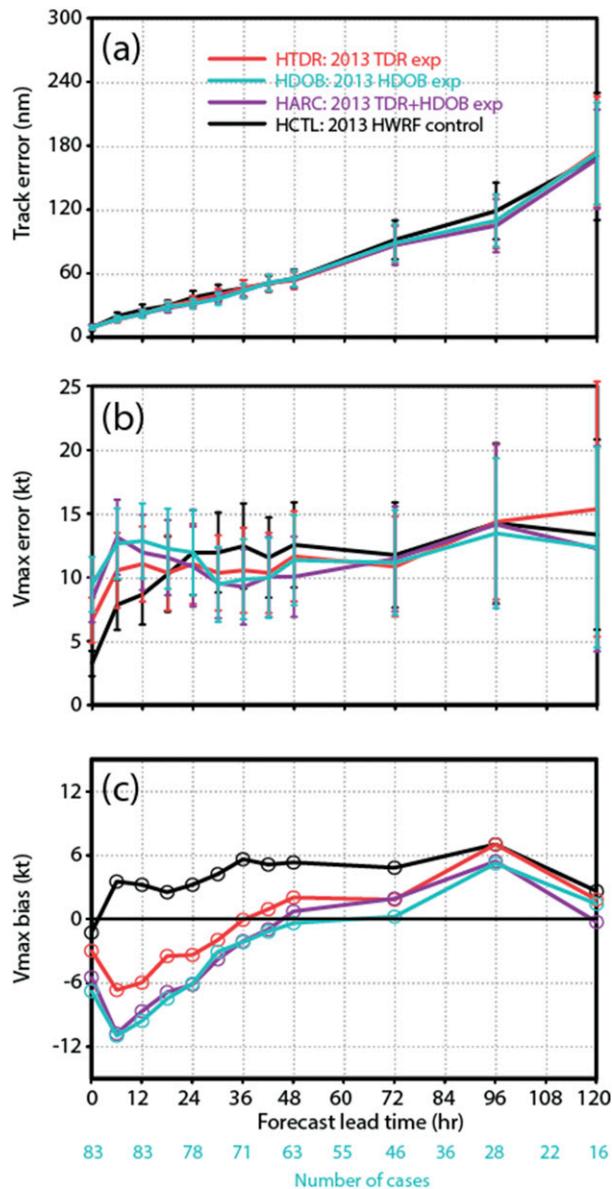


FIG. 11. As in the left column of Fig. 3, but for cycles when both HDOB and TDR data are available.

$V_{\max}$  at this forecast time is 87 kt, whereas in the best track, it is 115 kt. As a result of trying to match  $V_{\max}$ , VI at 0600 UTC 19 August strengthens the entire vortex even more (Fig. 15c). Vukicevic et al. (2013) provide a more quantitative assessment of the impacts of HWRf physics on spindown, as is discussed in section 4.

Finally, another major contributor to spindown in H213 is inevitably suboptimal forecast error covariance used during data assimilation. As described in section 2, H213 uses the GDAS perturbations for flow-dependent covariance in GSI. While this certainly is an improvement upon static error covariance, it is still inadequate

for the tropical cyclone inner core. Though it is beyond the scope of this paper to quantitatively assess the impacts of this issue in the experiments at hand, we can turn to recent results from others to assess its scope. In particular, Lu et al. (2017a) showed that a continuously cycled hybrid EnKF-Var system for HWRf (2014 version) can produce better analyses and forecasts by using a HWRf EnKF ensemble in place of the GFS ensemble. Lu et al. (2017b) used an upgraded version of this system to examine the entire life cycle of Hurricane Edouard (2014) and found it to significantly reduce short-term negative intensity bias, as compared with the operational HWRf. Further improvements were seen with an upgrade from 3DEnVar to 4DEnVar.

#### 4. Conclusions and discussion

This study assesses the performance of the data assimilation and modeling system that was implemented for the FY2013 upgrade of the HWRf Model. The GSI-based hybrid data assimilation system uses perturbations from the NCEP GFS GSI-EnKF hybrid system to estimate flow-dependent background error covariance, and data assimilation follows the vortex relocation and correction processes to enhance the representation of the TC inner core. The HWRf data assimilation system is examined in a series of experiments that assimilate different combinations of high-resolution, inner-core aircraft reconnaissance observations collected during the 2008–12 hurricane seasons. The impact of assimilating reconnaissance observations on the analysis and forecast of TCs is evaluated by comparing the track and intensity forecast errors in inner-core data assimilation experiments with the control experiment in which VI is applied without any inner-core data assimilation.

Our experiments with different combinations of inner-core data show that inner-core assimilation can be beneficial to H213, but its impacts depend strongly upon the storm initial intensity. Assimilating the TDR data have neutral to positive impact on TC track forecasts and a positive impact on TC intensity forecasts after the 18-h forecast lead time. The positive impact is mainly for weaker storms for which assimilation of the TDR data reduces the track errors and the positive intensity bias. Meanwhile, short-term intensity forecasts tend to be degraded, mostly due to spindown of major hurricane cases during the first 6 h. Assimilation of the flight-level and SFMR data improves the TC track forecasts and has neutral to positive impact on intensity forecast beyond 30-h forecast lead time. However, there is degradation of the short-term intensity forecast that is more severe in stronger storms. In those cases, forecast spindown can increase intensity

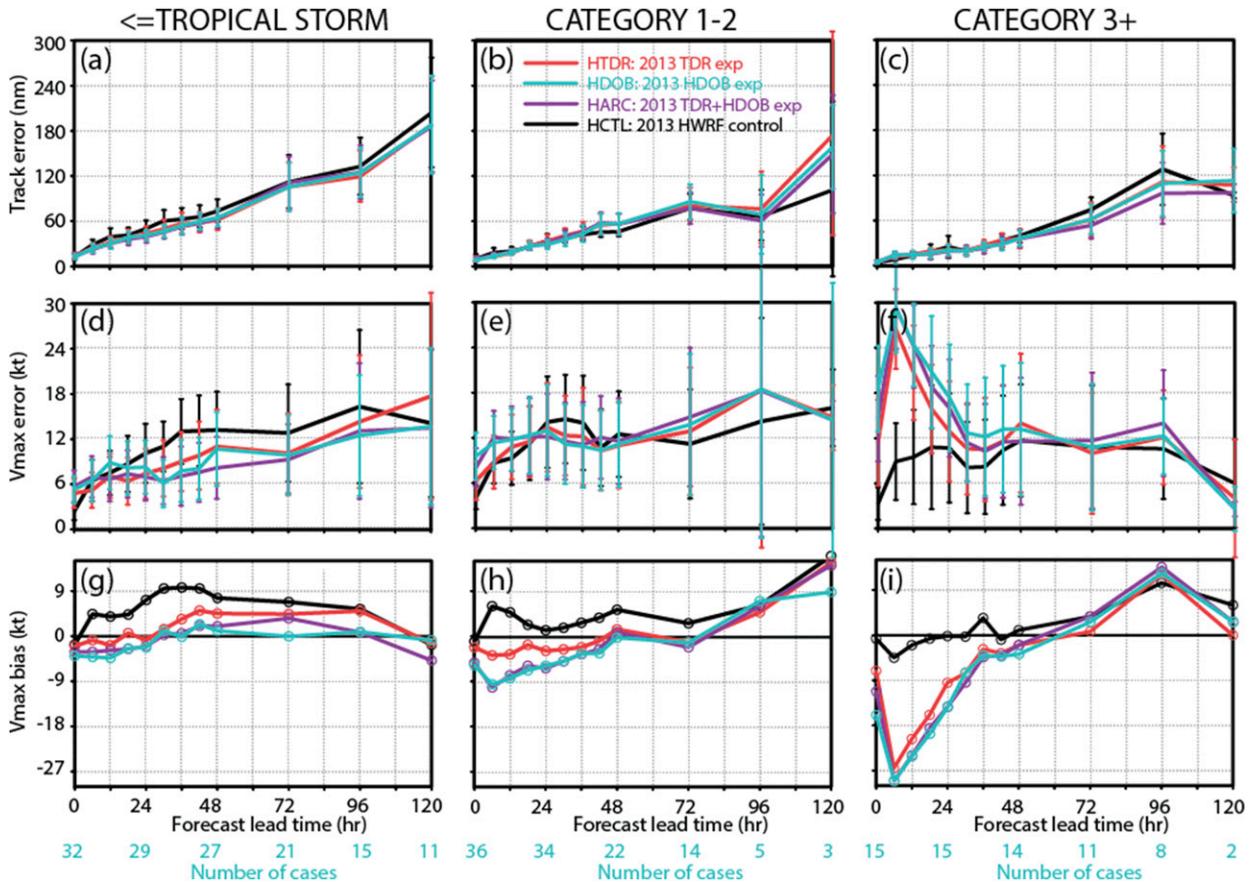


FIG. 12. As in Fig. 4, but for cycles with both TDR and HDOB data available.

forecast error for up to 48 h. Finally, assimilating all reconnaissance observations including TDR data produces similar or slightly improved track and intensity forecasts, compared with assimilating only flight-level, SFMR, and dropsonde observations.

There are both similarities and differences between our results and others that have looked at the impact of reconnaissance on tropical cyclone forecasts. Our results are most consistent with WZ16, who provide the most comprehensive assessment to date of the impact of reconnaissance (i.e., HDOB and TDR) on tropical cyclone forecasts. The results from that study showed that reconnaissance assimilation degraded Vmax forecasts during the first 24 h and improved them thereafter to 108 h, with peak improvement from about 36 to 66 h. This is a qualitatively similar result to the results obtained here. The track forecasts in WZ16 improved noticeably due to assimilation after about 36 h, with peak improvement between 60 and 84 h. Assimilation of inner-core data here improved track somewhat later in the forecast, and the maximum improvement was from 84 to 120 h.

The only other study to cleanly examine impact of inner-core reconnaissance over more than just a few cases was A15, and their results were somewhat different. In particular, Vmax improved through the duration of most of their forecasts, with the improvement of early lead times being due to TDR assimilation and the improvement of longer lead times being due to other reconnaissance. The improved performance at early lead times in A15 was likely a reflection of two or three factors. First, A15 only examined cases where TDR was available (note that the TDR data sample used in A15 was largely the same as that used in our study), whereas the TDR samples were a vast minority of WZ16 cases. In fact, the WZ16 results were qualitatively similar to the A15 experiment that did not assimilate TDR data. The second factor leading to improved performance in A15 during the first 24 h, as compared with our results, is that A15 used an EnKF that is able to generate mesoscale error covariance, whereas the covariance in H213 came from GDAS (this is discussed further below). Finally, in the experiment assimilating only HDOB data, A15 used a very different data sample, which can impact results.

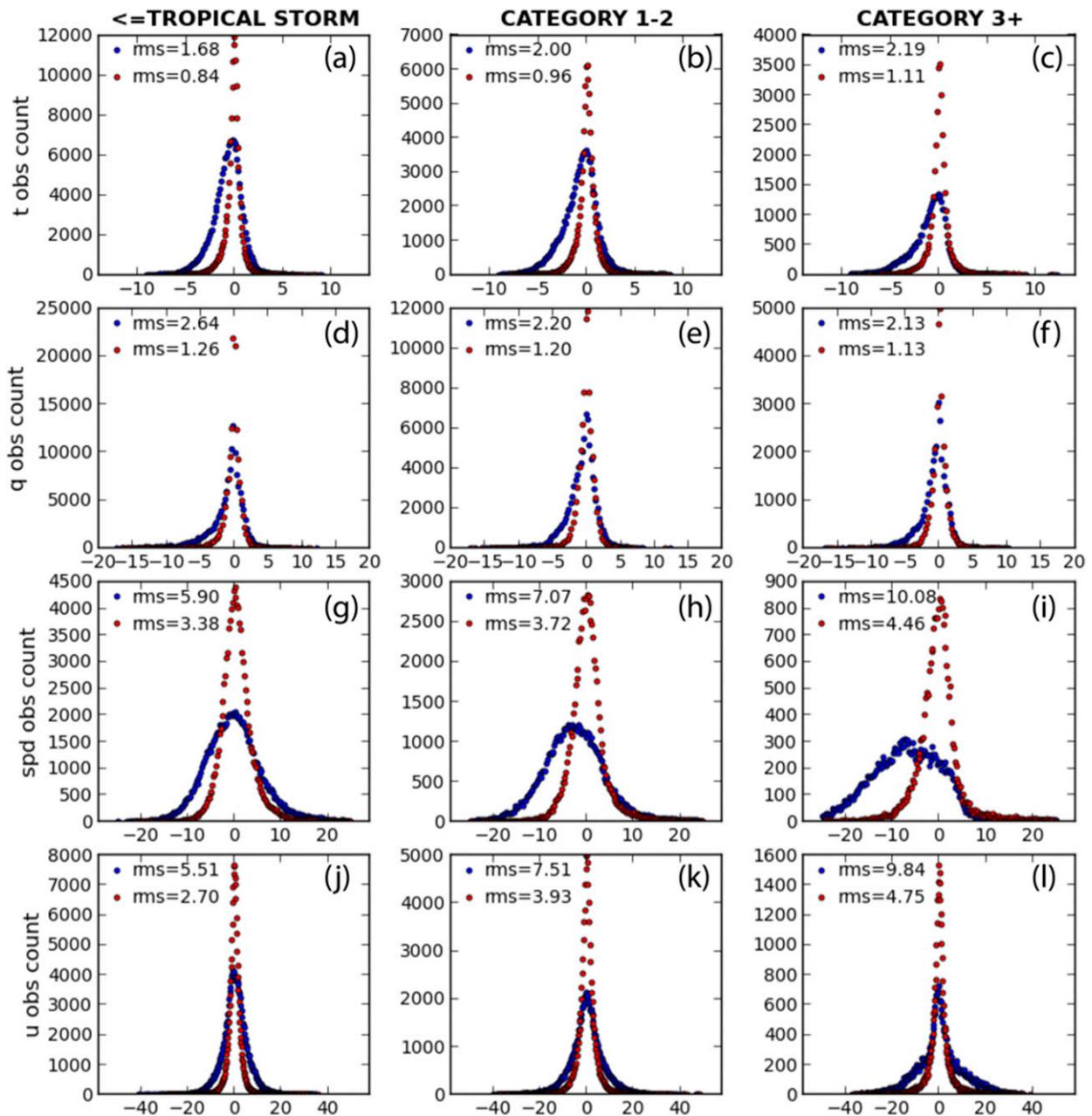


FIG. 13. Histograms of O-F (blue) and O-A (red) for flight-level (a)–(c) temperature (K), (d)–(f) specific humidity ( $\text{g kg}^{-1}$ ), (g)–(i) SFMR retrieved wind speed ( $\text{m s}^{-1}$ ), and (j)–(l) flight-level zonal wind ( $\text{m s}^{-1}$ ), grouped by storm initial intensity. The data are from all HDOB cases.

Analysis of the different experiments here reveals that the inner-core data assimilation benefits weaker storms for H213 because it offsets an intrinsic positive intensity bias that occurs in forecasts initialized only with VI (HCTL). Our analysis reveals that the bias is a result of TC vortices in H213 having stronger, more symmetric circulations when no inner-core observations are assimilated. Such overestimation of the TC intensity occurs partly because of a combination of model internal

biases and the fact that VI alters the wind distribution throughout the vortex, which is proportional to the difference between the forecasted and observed Vmax. The TCvitals information used to perform size and intensity corrections in VI (RMW, R34, ROCI, and Vmax) are only surface kinematic values. They cannot represent a fully three-dimensional kinematic and thermodynamic structure of the vortex. This is most problematic when asymmetries related to vortex tilt and

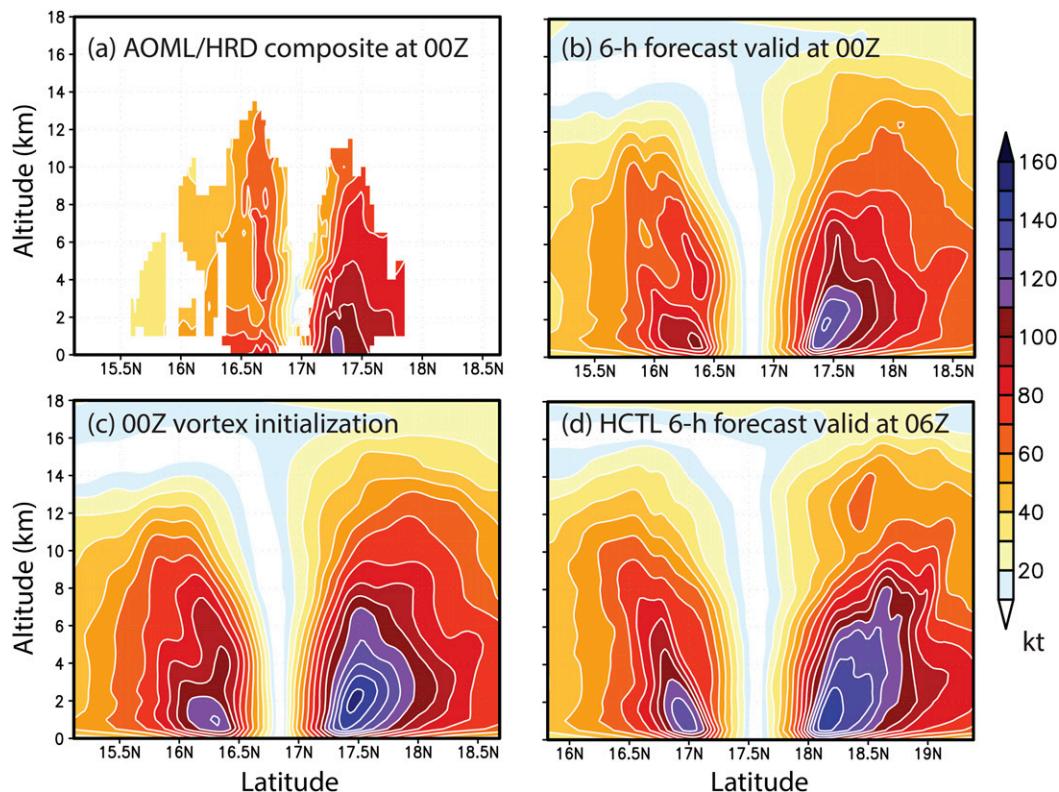


FIG. 14. South–north vertical cross-sections of zonal wind speed from (a) the HRD TDR analysis at 0000 UTC 19 Aug 2009; (b) the 6-h HCTL forecast valid at 0000 UTC; (c) vortex initialization at 0000 UTC; and (d) 6-h forecast from (c). Note that (b)–(d) are all from HCTL experiment. There is no inner-core DA in HCTL; the vortex is initialized through vortex initialization. Therefore, (c) is the initial condition used for the forecast in (d).

strong vertical wind shear are present, which frequently occurs in weaker storms. As such, inner-core data assimilation can provide more accurate structure of asymmetric systems that VI cannot adequately address. As long as the bias in the first guess is small, which is the case for weaker storms, moderate correction to vortex kinematic and thermodynamic structure through data assimilation can help improve forecast skill.

For strong storms, model biases appear to worsen, and the conflict between the vortex structure preferred in H213 and what actually occurs dominates any benefit from data assimilation. By adjusting the entire vortex to match the model  $V_{max}$  with that from TCVitals, VI reinforces aforementioned model biases to maintain a “fat and strong” inner-core structure for strong storms. Though assimilating inner-core data improves analyses in H213, severe problems arise because H213 is unable to produce a vortex with the proper structure. In particular, the resulting  $V_{max}$  has a negative bias, and the structure imparted by the observations is inconsistent with that preferred by the model physics. After the inevitable short-term spindown, VI resurrects the “fat and

strong” inner-core structure in its attempt to match the TCVitals during the next analysis cycle.

Though VI is still used in the current operational HWRF, by design, it cannot fully use information from different observations. Therefore, the opportunity for VI to improve is limited. It is expected that as model physics and the data assimilation system improve, VI will be phased out. More comprehensive experiments will be conducted to evaluate the role of VI as the modeling and data assimilation system evolves in the future.

Though this study is somewhat subjective in terms of assessing reasons for spindown, substantial insight can be gained by comparing these results with those from other DA systems that use the HWRF Model. In particular, the HEDAS system at NOAA HRD utilizes an EnKF with cycled covariance to perform advanced assimilation of a substantial number of tropical cyclones of varying intensity (Aksoy et al. 2013; A15). Furthermore, Lu et al. (2017a) developed a continuously cycled 3D-EnVar system for HWRF, and Lu et al. (2017b) conducted a thorough analysis of that system, as well as an upgrade to 4D-EnVar.

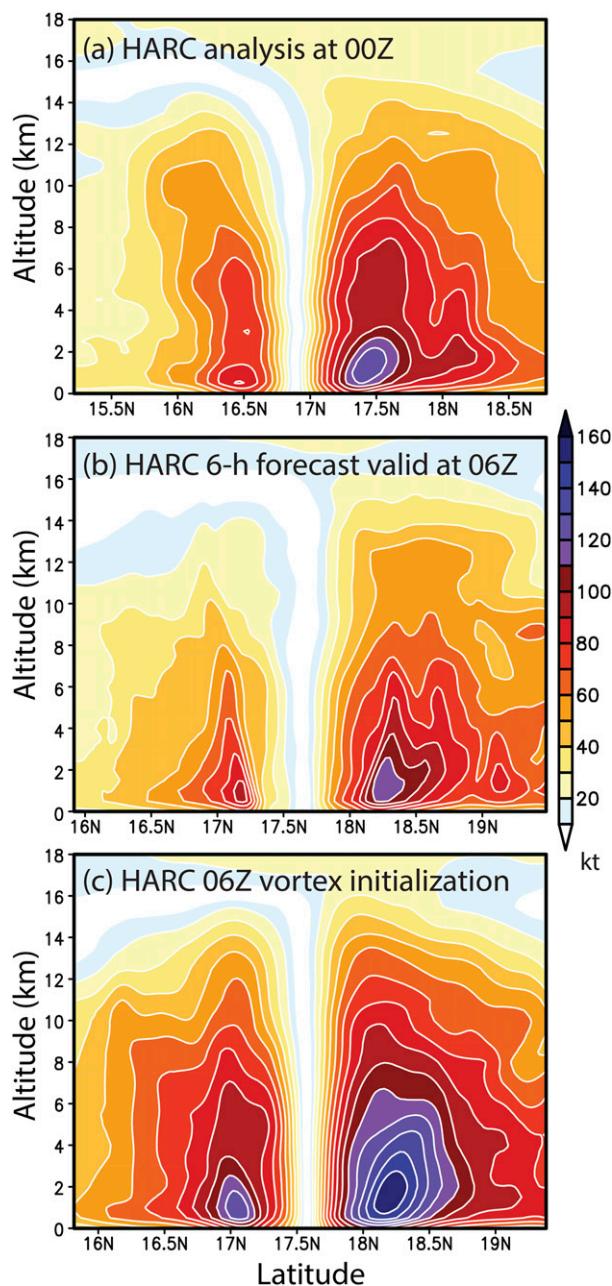


FIG. 15. As in Fig. 14, but for HARC (a) analysis at 0000 UTC 19 Aug 2009, (b) the 6-h forecast from (a), and (c) vortex initialization at 0600 UTC 19 Aug 2009.

Suboptimal error covariance undoubtedly contributes to spindown. The system developed by Lu et al. (2017b) examined the entire life cycle of Hurricane Edouard (2014) and found it to significantly reduce short-term negative intensity bias, as compared with the operational HWRf. Further improvements were seen with an upgrade from 3DEnVar to 4DEnVar. A similar 3DEnVar system was implemented in the 2017 operational HWRf with a primary benefit of reducing short-term negative

intensity bias, though it does not solve spindown for strong storms (Sippel et al. 2017). A case study by Lu et al. (2017b) shows that further upgrades to the system from 3DEnVar to 4DEnVar can improve intensity forecasts for earlier lead times.

Spindown issues are not limited to the operational HWRf. The stronger storms in A15, which used the more advanced HWRf data assimilation system in HEDAS, were also subject to substantial spindown, and Vukicevic et al. (2013) provided detailed analysis of the issue. That study revealed that HEDAS tends to exhibit a weak bias in the secondary circulations of strong hurricanes, such that the strong primary circulation cannot be supported. A model bias in PBL height, as well the lack of analysis for vertical velocity (forecasts in HEDAS started with zero vertical velocity at the time, and they still do in HWRf), are possible reasons for this bias. Since the HEDAS system uses the operational HWRf Model (albeit usually an upgrade or two behind), the operational HWRf also suffers the same model bias issues. Finally, even the advanced system in Lu et al. (2017b) can suffer spindown at times. Ongoing work has shown that physics changes can significantly reduce spindown in that system (Wang 2017), though a major challenge is to implement changes that yield appropriate inner-core structures but do not cause large positive intensity biases for weaker systems.

Some of the problems identified in Vukicevic et al. (2013) have also been identified and addressed in other work. For example, Zhang et al. (2015) and some earlier studies demonstrated the important role of the boundary layer on hurricane structure and intensity forecasts. Particularly, the vertical diffusion in the boundary layer has profound impact on controlling inner-core size, inflow strength and depth, and the strength and height of the maximum tangential wind speed. Improvement in vertical diffusion can improve storm structure with respect to aircraft observations. Recent work by Zhang and Marks (2015) also found that storm size and kinematic boundary height are sensitive to the horizontal diffusion. Lowering the surface drag coefficient in hurricanes, which was done starting with the 2016 implementation of HWRf, also significantly improves vertical wind profiles. Future improvements to HWRf will involve collaboration with scientists in the research community, such as HRD, to more systematically evaluate model bias and improve model physics.

The results here motivated further testing and implementation of inner-core reconnaissance assimilation in the operational HWRf system. In particular, assimilation of HDOB data began in the 2017 HWRf implementation (Sippel et al. 2017) after retrospectives

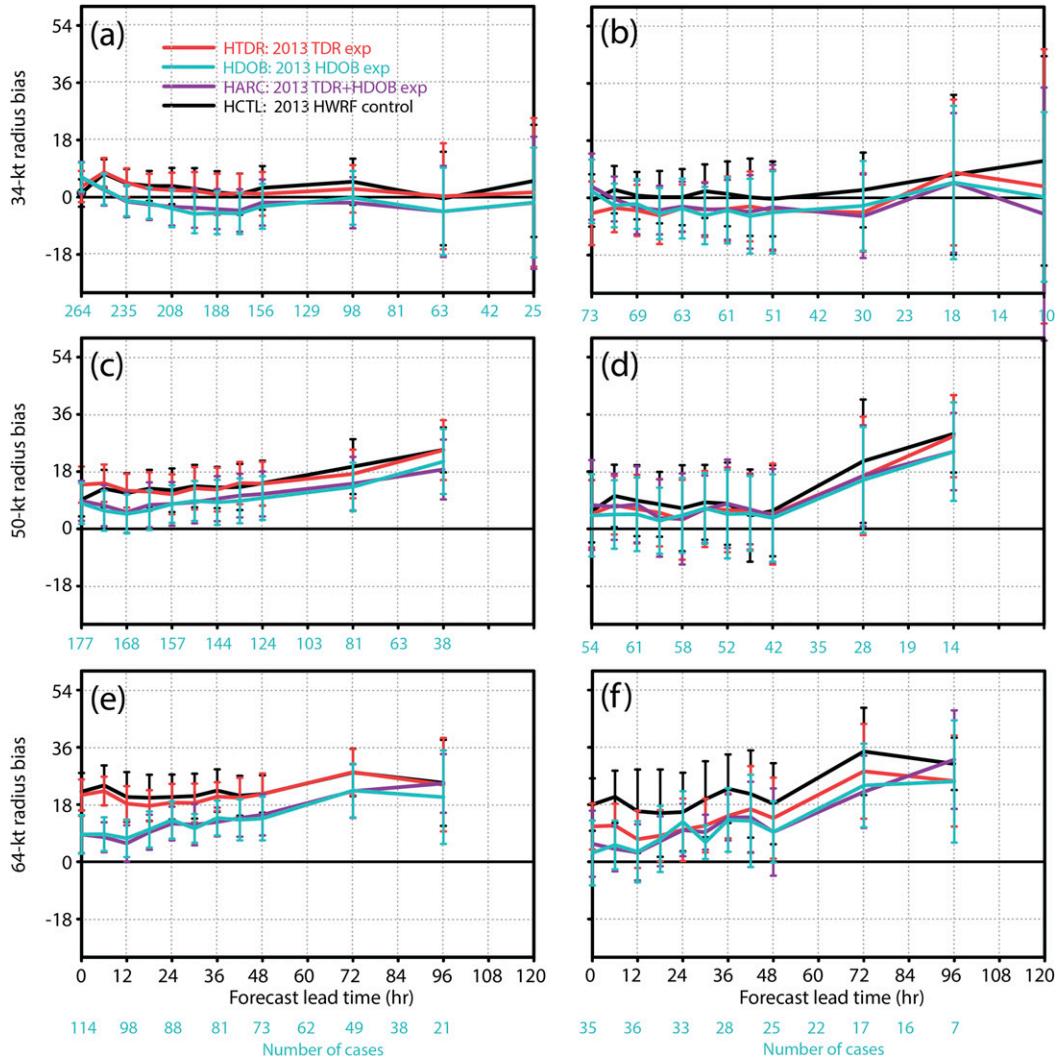


FIG. 16. Average wind radius bias (mean differences relative to the best track computed over all quadrants) are evaluated for: (a),(b) 34-; and (c),(d) 50- and (e),(f) 64-kt radii. (left) A homogeneous comparison of the four experiments for all cycles, as in the left column of Fig. 3; (right) a comparison for the cycle when both TDR and HDOB data are available, as in Fig. 11.

continued to show a considerable forecast benefit (operational assimilation of TDR data began in 2013, before this study was completed). Because of problems with spindown for strong storms, the current practice in the operational HWRF is to not use inner-core analysis increments for hurricanes (and instead only use VI; note that this special treatment for the inner-core analysis increments is not used in the experiments presented in this manuscript). As a result of these implementations, unpublished tests by NCEP have continually shown that inner-core reconnaissance has a meaningful positive impact on tropical cyclone track and intensity forecasts from HWRF. It is anticipated that GSI increments will be used for stronger storms as data assimilation advances, which should increase the benefit of inner-core data.

The above results have motivated significant improvements to HWRF data assimilation in the near future. The benefit of using more frequent cycling including cloud, precipitation, and vertical velocity analysis, as well as more advanced schemes, such as 4D-EnVar, will be further explored. In addition, an initialization scheme using incremental analysis updates (IAUs) (Bloom et al. 1996), which aims to improve the balance in the initial condition, is under development and will potentially be implemented soon. Besides data assimilation system upgrades, effort is being invested to improve the use of the aircraft observations through better quality control, observation error tuning, and ingesting more aircraft observations (e.g., observations from the NOAA Gulfstream IV and NASA Global Hawk aircraft), as well

as satellite observations. Finally, in addition to addressing the key issues of the data assimilation, NCEP is addressing the role of and possible alternatives to VI.

*Acknowledgments.* This work was supported by the NOAA Hurricane Forecast Improvement Project (HFIP). We acknowledge the NCEP/EMC data assimilation team, in particular, David Parrish, Wan-shu Wu, Daryl Kleist, Shun Liu, and John Derber, for their efforts on developing hybrid data assimilation scheme in GSI and sharing thoughts and techniques on improving data assimilation for HWRF. We thank John F. Gamache, Sim Aberson, Paul Reasor, and Robert Rogers for providing technical information about TDR data and TDR composites. We thank Fuqing Zhang, Altug Aksoy, and Xuguang Wang for sharing ideas and techniques on HWRF data assimilation. Finally, we acknowledge NOAA AOC, AOML, and the U. S. Air Force for their tremendous efforts to fly into hurricanes and collect the data used in this study and in operational HWRF.

#### REFERENCES

- Aberson, S. D., A. Aksoy, K. J. Sellwood, T. Vukicevic, and X. Zhang, 2015: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using HEDAS: Evaluation of 2008–11 HWRF forecasts. *Mon. Wea. Rev.*, **143**, 511–523, <https://doi.org/10.1175/MWR-D-14-00138.1>.
- , K. J. Sellwood, and P. A. Leighton, 2017: Calculating dropwindsonde location and time from TEMP-DROP messages for accurate assimilation and analysis. *J. Atmos. Oceanic Technol.*, **34**, 1673–1678, <https://doi.org/10.1175/JTECH-D-17-0023.1>.
- Aksoy, A., S. D. Aberson, T. Vukicevic, K. J. Sellwood, S. Lorsolo, and X. Zhang, 2013: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD's HEDAS: Evaluation of the 2008–11 vortex-scale analyses. *Mon. Wea. Rev.*, **141**, 1842–1865, <https://doi.org/10.1175/MWR-D-12-00194.1>.
- Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271, [https://doi.org/10.1175/1520-0493\(1996\)124<1256:DAUIAU>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2).
- Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043, <https://doi.org/10.1256/qj.04.15>.
- Christophersen, H., A. Aksoy, J. Dunion, and K. Sellwood, 2017: The impact of NASA Global Hawk unmanned aircraft dropwindsonde observations on tropical cyclone track, intensity, and structure: Case studies. *Mon. Wea. Rev.*, **145**, 1817–1830, <https://doi.org/10.1175/MWR-D-16-0332.1>.
- Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.*, **51**, 249–280, [https://doi.org/10.1175/1520-0469\(1994\)051<0249:ADMMPF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<0249:ADMMPF>2.0.CO;2).
- Franklin, J. L., M. L. Black, and K. Valde, 2003: GPS dropwindsonde wind profiles in hurricanes and their operational implications. *Wea. Forecasting*, **18**, 32–44, [https://doi.org/10.1175/1520-0434\(2003\)018<0032:GDWPIH>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0032:GDWPIH>2.0.CO;2).
- Gamache, J. F., 1997: Evaluation of a fully three-dimensional variational Doppler analysis technique. Preprints, *28th Conf. on Radar Meteorology*, Austin, TX, Amer. Meteor. Soc., 422–423.
- , 2005: Real-time dissemination of hurricane wind fields determined from airborne Doppler radar data. NOAA Final Rep. on JHT project, 38 pp., [http://www.nhc.noaa.gov/jht/2003-2005reports/DOPLRgamache\\_JHTfinalreport.pdf](http://www.nhc.noaa.gov/jht/2003-2005reports/DOPLRgamache_JHTfinalreport.pdf).
- Gopalakrishnan, S. G., F. Marks Jr., J. A. Zhang, X. Zhang, J.-W. Bao, and V. Tallapragada, 2013: A study of the impacts of vertical diffusion on the structure and intensity of the tropical cyclones using the high-resolution HWRF system. *J. Atmos. Sci.*, **70**, 524–541, <https://doi.org/10.1175/JAS-D-11-0340.1>.
- Hamill, T. M., J. S. Whitaker, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011: Predictions of 2010's tropical cyclones using the GFS and ensemble-based data assimilation methods. *Mon. Wea. Rev.*, **139**, 3243–3247, <https://doi.org/10.1175/MWR-D-11-00079.1>.
- Han, J., and H.-L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Wea. Forecasting*, **26**, 520–533, <https://doi.org/10.1175/WAF-D-10-05038.1>.
- Hendricks, E. A., M. S. Peng, X. Ge, and T. Li, 2011: Performance of a dynamic initialization scheme in the Coupled Ocean–Atmosphere Mesoscale Prediction System for tropical cyclones (COAMPS-TC). *Wea. Forecasting*, **26**, 650–663, <https://doi.org/10.1175/WAF-D-10-05051.1>.
- Janjic, Z., R. Gall, and M. E. Pyle, 2010: Scientific documentation for the NMM solver. NCAR Tech. Note NCAR/TN-477+STR, 53 pp., [https://dtcenter.org/HurrWRF/users/docs/scientific\\_documents/NMM\\_scientific\\_2-2-10\\_final.pdf](https://dtcenter.org/HurrWRF/users/docs/scientific_documents/NMM_scientific_2-2-10_final.pdf).
- Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433–451, <https://doi.org/10.1175/MWR-D-13-00351.1>.
- Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, <https://doi.org/10.1175/MWR-D-12-00254.1>.
- Lawless, A. S., 2010: A note on the analysis error associated with 3D-FGAT. *Quart. J. Roy. Meteor. Soc.*, **136**, 1094–1098, <https://doi.org/10.1002/qj.619>.
- Lee, W.-C., F. D. Marks, and C. Walther, 2003: Airborne Doppler radar data analysis workshop. *Bull. Amer. Meteor. Soc.*, **84**, 1063–1075, <https://doi.org/10.1175/BAMS-84-8-1063>.
- Li, Y., X. Wang, and M. Xue, 2012: Assimilation of radar radial velocity data with the WRF hybrid ensemble–3DVAR system for the prediction of Hurricane Ike (2008). *Mon. Wea. Rev.*, **140**, 3507–3524, <https://doi.org/10.1175/MWR-D-12-00043.1>.
- Liu, Q., X. Zhang, S. Trahan, and V. Tallapragada, 2012: Extending operational HWRF initialization to triple-nest HWRF system. *30th Conf. on Hurricanes and Tropical Meteorology*, Ponte Vedra Beach, FL, Amer. Meteor. Soc., 14A.6, <https://ams.confex.com/ams/30Hurricane/webprogram/Paper204853.html>.
- Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, <https://doi.org/10.1256/qj.02.132>.
- Lorsolo, S., J. Gamache, and A. Aksoy, 2013: Evaluation of the Hurricane Research Division Doppler radar analysis software using synthetic data. *J. Atmos. Oceanic Technol.*, **30**, 1055–1071, <https://doi.org/10.1175/JTECH-D-12-00161.1>.
- Lu, X., X. Wang, Y. Li, M. Tong, and X. Ma, 2017a: GSI-based ensemble-variational hybrid data assimilation for HWRF for hurricane initialization and prediction: Impact of various error

- covariances for airborne radar observation assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 223–239, <https://doi.org/10.1002/qj.2914>.
- , —, M. Tong, and V. Tallapragada, 2017b: GSI-based, continuously cycled, dual-resolution hybrid ensemble–variational data assimilation system for HWRf: System description and experiments with Edouard (2014). *Mon. Wea. Rev.*, **145**, 4877–4898, <https://doi.org/10.1175/MWR-D-17-0068.1>.
- Mellor, G. L., 2004: Users guide for a three-dimensional, primitive equation, numerical ocean model (June 2004 version). Princeton University Tech. Rep., 56 pp., [http://jes.apl.washington.edu/modsims\\_two/usersguide0604.pdf](http://jes.apl.washington.edu/modsims_two/usersguide0604.pdf).
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763, [https://doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCSS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2).
- Rabier, F., J.-N. Thépaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1861–1887, <https://doi.org/10.1002/qj.49712455005>.
- Sippel, J. A., M. Tong, and A. Mehra, 2017: 2016–2017 HWRf DA testing & plans. *71st Interdepartmental Hurricane Conf.*, Miami, FL, Office of the Federal Coordinator for Meteorology, [https://www.ofcm.gov/meetings/TCORF/ihc17/Session\\_07/7-3%20sippelnew\\_web.pdf](https://www.ofcm.gov/meetings/TCORF/ihc17/Session_07/7-3%20sippelnew_web.pdf).
- Tallapragada, V., and Coauthors, 2013: Hurricane Weather Research and Forecasting (HWRf) Model: 2013 scientific documentation. NCAR Development Testbed Center Rep., 99 pp., [http://www.dtcenter.org/HurrWRF/users/docs/scientific\\_documents/HWRfV3.5a\\_ScientificDoc.pdf](http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRfV3.5a_ScientificDoc.pdf).
- Trahan, S., and L. Sparling, 2012: An analysis of NCEP Tropical Cyclone Vitals and potential effects on forecasting models. *Wea. Forecasting*, **27**, 744–756, <https://doi.org/10.1175/WAF-D-11-00063.1>.
- Uhlhorn, E. W., P. G. Black, J. L. Franklin, M. Goodberlet, J. Carswell, and A. S. Goldstein, 2007: Hurricane surface wind measurements from an operational stepped frequency microwave radiometer. *Mon. Wea. Rev.*, **135**, 3070–3085, <https://doi.org/10.1175/MWR3454.1>.
- Vukicevic, T., A. Aksoy, P. Reasor, S. D. Aberson, K. J. Sellwood, and F. Marks, 2013: Joint impact of forecast tendency and state error biases in ensemble Kalman filter data assimilation of inner-core tropical cyclone observations. *Mon. Wea. Rev.*, **141**, 2992–3006, <https://doi.org/10.1175/MWR-D-12-00211.1>.
- Wang, X., 2010: Incorporating ensemble covariance in the grid-point statistical interpolation (GSI) variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, <https://doi.org/10.1175/2010MWR3245.1>.
- , 2017: DA/initialization team milestones and priorities. *2017 HFIP Annual Review Meeting*, Miami, FL, NOAA Hurricane Forecast Improvement Project, 26 pp., [http://www.hfip.org/events/annual\\_meeting\\_nov\\_2017/docs/Day-1\\_1415PM\\_XuguangWangDA-EnsDevelopment-v2.pdf](http://www.hfip.org/events/annual_meeting_nov_2017/docs/Day-1_1415PM_XuguangWangDA-EnsDevelopment-v2.pdf).
- , D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, <https://doi.org/10.1175/MWR-D-12-00141.1>.
- Weng, Y., and F. Zhang, 2012: Assimilating airborne Doppler radar observations with an ensemble Kalman filter for convection-permitting hurricane initialization and prediction: Katrina (2005). *Mon. Wea. Rev.*, **140**, 841–859, <https://doi.org/10.1175/2011MWR3602.1>.
- , and —, 2016: Advances in convection-permitting tropical cyclone analysis and prediction through EnKF assimilation of reconnaissance aircraft observations. *J. Meteor. Soc. Japan*, **94**, 345–358, <https://doi.org/10.2151/jmsj.2016-018>.
- Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916, [https://doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2).
- Zhang, F., and Y. Weng, 2015: Predicting hurricane intensity and associated hazards: A five-year real-time forecast experiment with assimilation of airborne Doppler radar observations. *Bull. Amer. Meteor. Soc.*, **96**, 25–33, <https://doi.org/10.1175/BAMS-D-13-00231.1>.
- , —, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125, <https://doi.org/10.1175/2009MWR2645.1>.
- , —, J. F. Gamache, and F. D. 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, <https://doi.org/10.1029/2011GL048469>.
- Zhang, J. A., and F. D. Marks, 2015: Effects of horizontal diffusion on tropical cyclone intensity change and structure in idealized three-dimensional numerical simulations. *Mon. Wea. Rev.*, **143**, 3981–3995, <https://doi.org/10.1175/MWR-D-14-00341.1>.
- , D. S. Nolan, R. F. Rogers, and V. Tallapragada, 2015: Evaluating the impact of improvements in the boundary layer parameterization on hurricane intensity and structure forecasts in HWRf. *Mon. Wea. Rev.*, **143**, 3136–3155, <https://doi.org/10.1175/MWR-D-14-00339.1>.
- Zhu, L., and Coauthors, 2016: Prediction and predictability of high-impact western Pacific landfalling Tropical Cyclone Vicente (2012) through convection-permitting ensemble assimilation of Doppler radar velocity. *Mon. Wea. Rev.*, **144**, 21–43, <https://doi.org/10.1175/MWR-D-14-00403.1>.