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

ALTUĞ AKSOY

Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and NOAA/AOML/Hurricane Research Division, Miami, Florida

SIM D. ABERSON AND TOMISLAVA VUKICEVIC

NOAA/AOML/Hurricane Research Division, Miami, Florida

KATHRYN J. SELLWOOD, SYLVIE LORSOLO, AND XUEJIN ZHANG

Cooperative Institute for Marine and Atmospheric Studies, University of Miami, and NOAA/AOML/Hurricane Research Division, Miami, Florida

(Manuscript received 10 July 2012, in final form 3 October 2012)

ABSTRACT

The Hurricane Weather Research and Forecasting (HWRF) Ensemble Data Assimilation System (HEDAS) is developed to assimilate tropical cyclone inner-core observations for high-resolution vortex initialization. It is based on a serial implementation of the square root ensemble Kalman filter (EnKF). In this study, HWRF is used in an experimental configuration with horizontal grid spacing of 9 (3) km on the outer (inner) domain. HEDAS is applied to 83 cases from years 2008 to 2011. With the exception of two Hurricane Hilary (2011) cases in the eastern North Pacific basin, all cases are observed in the Atlantic basin. Observed storm intensity for these cases ranges from tropical depression to category-4 hurricane.

Overall, it is found that high-resolution tropical cyclone observations, when assimilated with an advanced data assimilation technique such as the EnKF, result in analyses of the primary circulation that are realistic in terms of intensity, wavenumber-0 radial structure, as well as wavenumber-1 azimuthal structure. Representing the secondary circulation in the analyses is found to be more challenging with systematic errors in the magnitude and depth of the low-level radial inflow. This is believed to result from a model bias in the experimental HWRF caused by the overdiffusive nature of the planetary boundary layer parameterization utilized. Thermodynamic deviations from the observed structure are believed to be caused by both an imbalance between the number of the kinematic and thermodynamic observations in general and the suboptimal ensemble covariances between kinematic and thermodynamic fields. Future plans are discussed to address these challenges.

1. Introduction

Numerical prediction of tropical cyclones (TCs) continues to be a challenge. Although improvements in track forecasts have been relatively steady in recent years, virtually no improvement has been made in forecasting intensity (e.g., Berg and Avila 2011). Several factors could be contributing to this, including model deficiencies, suboptimal initialization caused by the lack of observations in the peripheral environment as well as the inner-core circulation of a TC, and how existing observations are incorporated into model initial conditions through data assimilation. The goal of the present study is to focus on the data assimilation aspects of the TC prediction problem by exploring the impact of observations on the analyses of high-resolution vortex structure using the ensemble Kalman filter (EnKF).

The EnKF is an advanced data assimilation technique that utilizes an ensemble of short-range forecasts to estimate flow-dependent spatial and cross correlations for data assimilation (Evensen 1994; Houtekamer and Mitchell

Corresponding author address: Dr. Altuğ Aksoy, CIMAS, Rosenstiel School of Marine and Atmospheric Science, 4600 Rickenbacker Cswy., Miami, FL 33149.
E-mail: aaksoy@rsmas.miami.edu

DOI: 10.1175/MWR-D-12-00194.1

1998). Recent success with assimilating radar observations of continental convective storms (Snyder and Zhang 2003; Zhang et al. 2004; Dowell et al. 2004; Dowell et al. 2011; Aksoy et al. 2009, 2010) has raised hopes that high-resolution TC models, too, can benefit from the use of the EnKF. In a proof-of-concept study, Zhang et al. (2009) demonstrated that a model initialized with the EnKF exhibited more skill than the same model initialized with three-dimensional variational data assimilation (3DVAR) in predicting the evolution of a landfalling hurricane using observations from a land-based radar. The same data assimilation system was then recently tested with airborne Doppler radar observations (Weng and Zhang 2012; F. Zhang et al. 2011) and demonstrated improvement in the representation of the vortex structure in Hurricane Katrina (2005), as well as smaller intensity forecast errors in 61 cases from 2008 to 2010 when compared to operational dynamical models. In this study, the EnKF is combined with the National Oceanic and Atmospheric Administration's (NOAA) Hurricane Weather Research and Forecasting (HWRF; Gopalakrishnan et al. 2010) model to obtain high-resolution TC vortex analyses.

The HWRF Ensemble Data Assimilation System (HEDAS; Aksoy et al. 2012, hereafter A12) is an ensemble-based data assimilation system developed at NOAA's Hurricane Research Division (HRD) of the Atlantic Oceanographic and Meteorological Laboratory (AOML) to utilize high-resolution TC observations collected by NOAA's WP-3D (P-3) aircraft (Aberson et al. 2006), high-altitude Gulfstream-IV (G-IV) jet (Aberson 2009), as well as the C-130 aircraft of the 53rd Weather Reconnaissance Squadron of the U.S. Air Force Reserve Command (AFRC; e.g., Rappaport et al. 2009). HEDAS combines an EnKF with HRD's experimental HWRF model (Gopalakrishnan et al. 2012; X. Zhang et al. 2011) to carry out short-range ensemble forecasts during the cycling of observations. A12 demonstrated the value of assimilating simulated airborne Doppler radar radial wind data with HEDAS by showing not only direct positive impact on the vortex wind structure but also indirect positive impact on the thermodynamic structure.

The current article focuses on the assimilation of *real* airborne TC observations using HEDAS. Data assimilation is carried out for 83 cases (20 individual TCs) spanning the 2008–11 hurricane seasons. The TC position, intensity, and structure in the high-resolution vortex-scale analyses are evaluated by comparison to observations and other analyses. To the authors' knowledge, this is the first comprehensive study to investigate statistically the impacts of assimilating high-resolution TC observations on the analysis of vortex structure by systematic comparison to observed structure. While the focus here is only on vortex analyses produced by HEDAS, a future study will investigate the characteristics of the forecasts initialized with these analyses.

The details of the data assimilation and modeling aspects are described in section 2. Section 3 explains the cases considered in the study. Section 4 continues with the presentation of results; the summary and discussion are in section 5.

2. The real-time HEDAS

a. HRD's experimental HWRF

Most of the differences between HRD's experimental HWRF and the NOAA/National Centers for Environmental Prediction (NCEP) operational HWRF arise from the choice of physical parameterization schemes and resolution. In the current study, the HRD experimental HWRF is configured with a pair of two-way-interacting computational domains (see Table 1 for details). The vortex-following nest motion of the inner domain (Gopalakrishnan et al. 2002, 2006) is suppressed during spinup and data assimilation cycles and all ensemble members are initialized with collocated inner domains to facilitate gridpoint-based spatial covariance computations in the EnKF. The $10° \times 10°$ inner nest is bigger than that of A12 and encompasses the entire circulation zones of storms with varying rates of forward motion. A thorough comparison of the physics parameterizations used in the experimental and operational versions of HWRF can be found in Gopalakrishnan et al. (2012).

b. HEDAS

HEDAS is based on a serial implementation of the square root EnKF of Whitaker and Hamill (2002). In a serial update loop, each observation is treated as a scalar quantity, and the update equations are simplified following Snyder and Zhang (2003). The horizontal length scale of covariance localization (Gaspari and Cohn 1999) is chosen so that most of the vortex is updated given the limited spatial distribution of observations in each cycle (see discussion in A12). Further technical details of HEDAS are explained in A12.

In its real-time application, HEDAS uses 30 ensemble members. The initial and lateral boundary ensemble perturbations are obtained from the experimental, EnKF-based global ensemble prediction system developed for the NCEP Global Forecast System (GFS). The details of this system and its performance for the prediction of 2009 and 2010 TCs are summarized in Hamill et al. (2011a,b). An ensemble spinup is initialized 6 h prior to the synoptic time around which a NOAA P-3 flight is

TABLE 1. Summary of the experimental setup.

Feature	Explanation
HEDAS	
Filter type	Ensemble square root filter
Ensemble size, initialization	30 members, initialized from GFS/EnKF analysis ensemble, 3–4-h spinup
Covariance localization, radii of influence	Through a compactly supported fifth-order correlation function, 80 grid distances in the horizontal and 15 model levels in the vertical
Assimilation domain	Static inner nest during spinup and assimilation, 3-km horizontal resolution, with a 10-gridpoint ''buffer zone'' to outer nest
Assimilation frequency, period	1 h, duration of availability of NOAA P-3 observations
Assimilated observation types	Doppler radial velocity superobservations, dropsonde and flight-level wind and temperature, and SFMR 10-m wind speed
Experimental HWRF	
Model version	WRF-Nonhydrostatic Mesoscale Model (NMM) core, version 3.0.1.1
Resolution	Horizontal, 9-km outer nest of approx size $60° \times 60°$ and 3-km inner nest of approx size $10° \times 10°$; vertical, 42 eta levels with model top at 50 hPa
Physics	Ferrier et al. (2002) microphysics, Hong and Pan (1998) cumulus (only in the outer domain, explicit convection in the inner domain), Dudhia (1989) shortwave radiation, Mlawer et al. (1997) longwave radiation, Hong and Pan (1996) PBL, Moon et al. (2007) surface layer, and Ek et al. (2003) land surface parameterization

centered. The spinup is carried out for 3–4 h (until the first observations are available) to develop the covariance structures relevant for the scales at which the data assimilation is performed.[1] The experimental GFS/EnKF initial perturbations, without any covariance inflation, are found to result in comparable ensemble spread to that in A12 [see section 5a(2) for more details]. Furthermore, A12 found that the impact of covariance inflation quickly became marginal as the rate of inflation was increased and suggested unrepresented model error to be an important cause of spread deficiency. Whitaker and Hamill (2012) discuss that it is generally more difficult to treat model-error-related spread deficiency either through inflation or physically based approaches. Given such inherent difficulties, the only marginal improvements beyond the current level of ensemble spread observed in A12 with inflation, and the ad hoc nature of inflation that renders difficult the tuning for systematic improvements, no covariance inflation is applied in the real-data experiments discussed here.

[1] While a longer spinup period could improve the background multivariate covariance structures (A12 found that a 12-h period was needed for the initial adjustment in a nature run at 1.5-km grid spacing), longer spinup periods would also result in greater deviations of the background state from observations and greater errors in the environmental flow, both of which would deteriorate the analysis quality. Indeed, our tests with a 6-h-longer spinup period did not produce analyses that were measurably better than those with the current spinup length of 3–4 h (not shown).

3. 2008–11 aircraft cases considered

In this study, data assimilation experiments are carried out only when Doppler radar observations were available from NOAA P-3 flights. A total of 83 such cases (20 TCs) are considered (see Table 2 for a list of the individual cases). All but two cases (both from 2011's Hurricane Hilary in the eastern North Pacific) were in the Atlantic basin. The geographical distribution of the observed positions for these cases, as obtained from the National Hurricane Center's (NHC's) hurricane database (HURDAT; Landsea et al. 2004), also known as the ''best track database,'' is shown in Fig. 1a. The general proximity of the cases to land is caused by the range limitations of the NOAA P-3 aircraft. The distribution of the cases according to their best-track intensity category is shown in Fig. 1b. A skewed distribution that peaks at tropical storm intensity (17.5–32 m s^{-1}) is evident. Overall, more than half of the cases are tropical storms or category-1 hurricanes (17.5–42 m s^{-1}).

4. Observations assimilated

The types of observations assimilated in HEDAS include Doppler radial wind superobservations (superobs; see A12 for details), GPS dropwindsonde [to be called dropsonde for brevity hereafter; Hock and Franklin (1999)] data, aircraft flight-level wind and temperature observations, and Stepped Frequency Microwave Radiometer (SFMR; Uhlhorn et al. 2007) wind speed retrievals at 10-m altitude. Further details on the observations and their processing can be found in Table 3. Observations are assimilated in 1-h assimilation windows according to

TABLE 2. Summary of the cases considered (1 kt = 0.5144 m s^{-1}).

Storm name[a]	Verification time and date[b]	Obs intensity[c] (kt)	Storm name[a]	Verification time and date[b]	Obs intensity[c] (kt)
Dolly	1200 UTC 20 Jul 2008	40	Alex	0000 UTC 29 Jun 2010	55
Dolly	0000 UTC 21 Jul 2008	45	TD2	0000 UTC 7 Jul 2010	N/A
Dolly	1200 UTC 21 Jul 2008	45	TD2	1200 UTC 7 Jul 2010	25
Dolly	0000 UTC 22 Jul 2008	45	TD2	0000 UTC 8 Jul 2010	30
Dolly	1200 UTC 22 Jul 2008	55	Earl	0000 UTC 29 Aug 2010	55
Fay	1200 UTC 14 Aug 2008	N/A	Earl	1200 UTC 29 Aug 2010	65
Fay	0000 UTC 15 Aug 2008	N/A	Earl	0000 UTC 30 Aug 2010	85
Fay	0600 UTC 15 Aug 2008	N/A	Earl	1200 UTC 30 Aug 2010	105
Fay	1800 UTC 15 Aug 2008	35	Earl	0000 UTC 31 Aug 2010	115
Fay	1800 UTC 18 Aug 2008	50	Earl	1200 UTC 1 Sep 2010	110
Fay	0600 UTC 19 Aug 2008	55	Earl	0000 UTC 2 Sep 2010	120
Gustav	0000 UTC 30 Aug 2008	75	Earl	1200 UTC 2 Sep 2010	115
Gustav	1200 UTC 30 Aug 2008	110	Earl	0000 UTC 3 Sep 2010	90
Gustav	0000 UTC 31 Aug 2008	120	Earl	1800 UTC 3 Sep 2010	75
Gustav	1200 UTC 31 Aug 2008	100	Earl	0000 UTC 4 Sep 2010	60
Gustav	0000 UTC 1 Sep 2008	95	Karl	0000 UTC 13 Sep 2010	N/A
Gustav	1200 UTC 1 Sep 2008	95	Karl	1200 UTC 13 Sep 2010	N/A
Ike	0000 UTC 10 Sep 2008	65	Karl	0000 UTC 14 Sep 2010	25
Ike	1200 UTC 10 Sep 2008	80	Karl	1800 UTC 16 Sep 2010	70
Ike	0000 UTC 11 Sep 2008	85	Richard	0600 UTC 23 Oct 2010	45
Ike	1200 UTC 11 Sep 2008	85	Tomas	0000 UTC 4 Nov 2010	40
Ike	0000 UTC 12 Sep 2008	85	Tomas	1200 UTC 4 Nov 2010	45
Ike	1800 UTC 12 Sep 2008	95	Tomas	0000 UTC 5 Nov 2010	60
Kyle	0000 UTC 23 Sep 2008	N/A	Tomas	1200 UTC 6 Nov 2010	60
Kyle	1200 UTC 24 Sep 2008	N/A	Tomas	0000 UTC 7 Nov 2010	70
Kyle	0000 UTC 25 Sep 2008	30	Irene	0000 UTC 24 Aug 2011	80
Kyle	1200 UTC 25 Sep 2008	40	Irene	1200 UTC 24 Aug 2011	105
Kyle	0000 UTC 26 Sep 2008	45	Irene	1200 UTC 25 Aug 2011	90
Kyle	1800 UTC 26 Sep 2008	50	Irene	0000 UTC 26 Aug 2011	90
Kyle	0000 UTC 27 Sep 2008	55	Irene	1200 UTC 26 Aug 2011	85
Kyle	1800 UTC 27 Sep 2008	65	Irene	0000 UTC 27 Aug 2011	75
Paloma	0600 UTC 7 Nov 2008	65	Irene	1200 UTC 27 Aug 2011	75
Paloma	1800 UTC 7 Nov 2008	80	Lee	0000 UTC 2 Sep 2011	30
Paloma	1800 UTC 8 Nov 2008	125	Ophelia	1800 UTC 24 Sep 2011	40
Ana	0000 UTC 17 Aug 2009	N/A	Hilary	1800 UTC 28 Sep 2011	70
Bill	0000 UTC 19 Aug 2009	105	Hilary	1800 UTC 29 Sep 2011	50
Bill	1200 UTC 19 Aug 2009	115	Rina	0000 UTC 26 Oct 2011	100
Bill	0000 UTC 20 Aug 2009	115	Rina	1800 UTC 26 Oct 2011	80
Bill	1200 UTC 20 Aug 2009	105	Rina	0000 UTC 27 Oct 2011	80
Danny	1200 UTC 26 Aug 2009	40	Rina	1800 UTC 27 Oct 2011	60
Danny	0000 UTC 27 Aug 2009	50			
Danny	1200 UTC 27 Aug 2009	50			
Danny	0000 UTC 28 Aug 2009	40			

[a] All cases except Hilary (2011, eastern Pacific basin) were observed in the Atlantic basin.
[b] Verification time here is defined as the closest synoptic time to the time of a final analysis.
[c] Storm intensity for cases that did not exist in the best-track database is denoted as N/A.

their sampling time. While the selection of cases is restricted to the availability of P-3 Doppler radar data, 1-h assimilation cycles in a given case are carried out for as long as the Doppler radar wind observations are available within the domain. The number of cases as a function of the number of assimilation cycles is shown in Fig. 2. The distribution peaks sharply at 5 cycles, which reflects the typical ~4–5-h duration that the NOAA P-3 aircraft remains ''on station'' within the storm circulation. Fewer assimilation cycles are the result of short on-station times for distant storms and/or aborted missions, while the greater number of cycles reflects longer data availability from overlaps with NOAA G-IV or AFRC C-130 flights and/or long on-station times for storms that are closer to deployment sites.

The distribution of the number of cases as a function of the number of observations assimilated by platform (Fig. 3, top histograms in each panel) reveals that the Doppler wind observations significantly outnumber observations from other platforms (by one order of magnitude

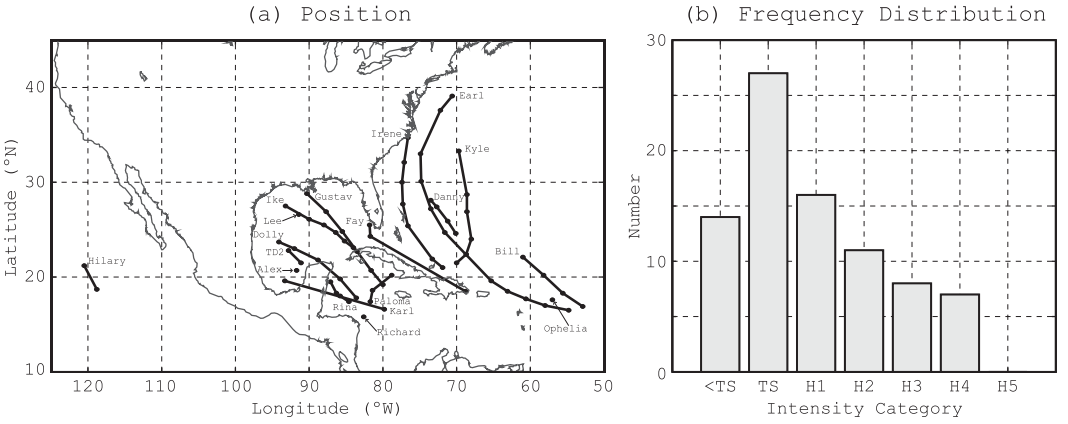

FIG. 1. (a) Geographical distribution of storm tracks and (b) frequency distribution of storm intensity for the 83 cases (20 TCs) examined in this study. Cases that do not exist in the best-track database (see Table 2) are assigned to the "<TS" category. TS and H stand for tropical storm and hurricane, respectively.

in the case of flight-level observations and by two orders of magnitude in the cases of dropsonde and SFMR observations). When broken down by intensity category (Fig. 3, 2D matrix plots in each panel), stronger TCs generally have more observations assimilated for each platform. The modes of these distributions also shift toward higher values for each platform, except for dropsonde observations that have the greatest number of cases in the smallest observation bin for all intensity categories. These differences can be explained by a combination of the following facts. 1) More Doppler wind observations are obtained in stronger storms as more symmetric convection results in greater areal and vertical hydrometeor coverage. 2) More flight-level and SFMR observations are obtained in stronger storms as the likelihood of overlapping NOAA P-3 and AFRC C-130 flights increases with storm intensity. 3) The number of dropsondes that are launched does not depend on storm intensity.

When the average number of observations assimilated per case is plotted by height (Fig. 4), the distribution of flight-level observations is broad (Fig. 4a) with secondary maxima at low altitudes for relatively weak TCs. This results from the combination of flight-level variability and the likelihood of low-level AFRC C-130 flights in weak TCs. In stronger TCs, the distributions become unimodal and peak near 3-km altitude, which is the standard AFRC C-130 and NOAA P-3 flight altitude [near 10 000 ft (3048 m)] in mature storms. A similar pattern of sharpening distributions with increasing intensity is also evident for Doppler wind observations (Fig. 4b), although the peaks are lower in altitude,

TABLE 3. Observation platforms/types assimilated.

Platform/observation type	Observation error (std dev)	Notes on processing
Tail Doppler radar (P-3)		
Radial wind speed	2 m s^{-1}	Processed into superobs[a]
HDOBS (P-3, C-130)		
Flight-level zonal and meridional wind speed	2 m s^{-1}	Observations with "questionable" quality control values removed
Flight-level temperature	0.5 K	Correction applied based on dewpoint temperature if necessary[b]
SFMR wind speed	Rain-rate dependent[c] with mean ~5 m s^{-1}	Observations with "questionable" quality control values removed
TEMPDROP (P-3, C-130, G-IV)		
Dropsonde zonal/meridional wind speed	2 m s^{-1}	Position and time interpolated based on drift deduced from wind speed, time, and location of release and splash
Dropsonde temperature	0.5 K	Position and time interpolated based on drift deduced from wind speed, time, and location of release and splash

[a] As in Aksoy et al. (2012).
[b] As in Eastin et al. (2002).
[c] As deduced from Uhlhorn et al. (2007).

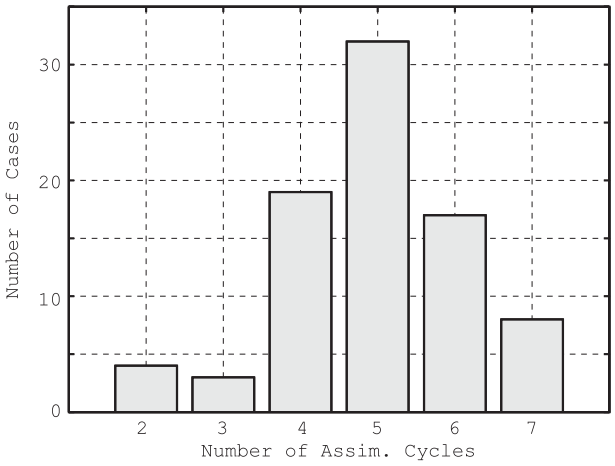

FIG. 2. Distribution of the number of cases with respect to the number of assimilation cycles.

indicating that such peaks occur below the aircraft. Finally, the peaks for the dropsonde observations also shift upward with intensity, which results from the generally lower NOAA P-3 and AFRC C-130 aircraft altitudes in weaker storms. Dropsonde observations sampled above approximately 3–4-km altitude are generally from environmental surveillance missions.

5. Results

a. Observation-space diagnostics

In this section, the focus is on innovation-based statistics, where innovations represent observation-minus-model (forecast or analysis) differences. The reader is reminded that in all experiments, all observation types are assimilated. Consequently, innovation statistics shown here for individual observation types still reflect impacts from other observation types.

First-cycle prior innovations reflect the observation-background differences before the onset of data assimilation; that is, they reflect the deviations from observations in the GFS/EnKF global initial conditions carried over into the HEDAS ensemble spinup. Last-cycle posterior innovations reflect the observation–analysis differences after the completion of assimilation of all inner-core data, where the memory of the global initial conditions is expected to be less impactful than the combined effects of the EnKF and the experimental HWRF through cycling. As the observations that are used in the first-cycle prior statistics are spatially displaced and 4–5 h apart from those in the last-cycle posterior statistics, the comparisons are believed to be sufficiently independent.

1) FREQUENCY DISTRIBUTIONS OF OBSERVATION INNOVATIONS

Figure 5 shows the frequency distributions of innovations for various observation types to depict the overall statistical behavior of the analyses in direct comparison to observations. The statistics are accumulated over all cases of interest to emphasize systematic patterns.

There appear to be no discernible biases associated with the model representation of Doppler wind observations. The main reason for this is that a Doppler wind observation's directional information is *relative* to the tail radar location within a given storm. Depending on the aircraft position, flight-track direction, and antenna scanning direction, a Doppler wind measurement is equally likely to be positive (away from the radar) or negative (toward the radar). Accumulated over sufficiently many observations over many cases, any apparent observation innovation biases therefore tend to cancel each other out so that innovation population density peaks at approximately 0 m s^{-1}. Meanwhile, the wide probability distribution of the first-cycle prior innovations clearly points to large wind errors carried over from the lower-resolution initial GFS/EnKF ensemble. The errors in the last-cycle posterior innovations appear to be considerably reduced, as deduced from the much narrower innovation distribution of Doppler wind observations. There is a 46% likelihood of encountering an absolute Doppler wind speed error of 3 m s^{-1} or less in the background field at the time of the first assimilation cycle. Through data assimilation, that likelihood is increased to 85% in the final analysis. Conversely, there is a 3% likelihood to encounter an absolute Doppler wind speed error of 20 m s^{-1} or more in the background field at the first assimilation cycle, while this likelihood is reduced to 0.01% by data assimilation in the final analysis.

SFMR observations carry information about the wind speed at 10 m and therefore are more likely to reflect model–analysis biases in storm intensity than are Doppler wind observations. Indeed, the probability distribution of the first-cycle prior SFMR innovations peaks between 1 and 3 m s^{-1}, indicating that the background 10-m wind speed is systematically underpredicted. An alternative way of looking at this is to measure the total probability of underprediction versus overprediction by the first-cycle background: the likelihood of encountering underpredicted 10-m wind speed is 68% as opposed to 32% for overprediction. Data assimilation helps to reduce this apparent bias and symmetrizes the posterior distribution of innovations in the final analysis: the likelihood of encountering underpredicted 10-m wind speed is now 57%, suggesting that observation–analysis differences

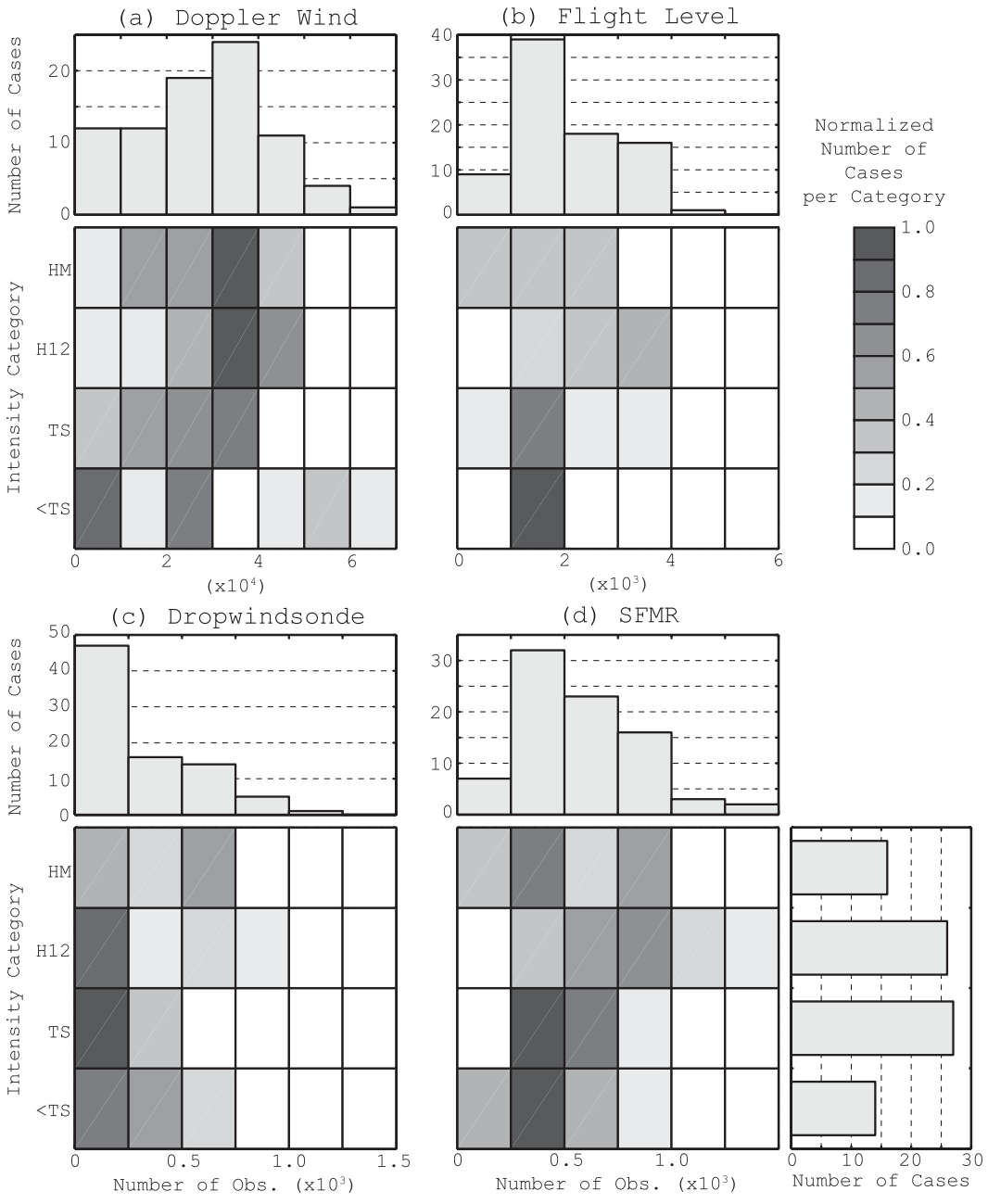

FIG. 3. Distribution of the number of cases according to the number of observations assimilated per platform and intensity category (2D matrix plots in each panel, gray scale at top right), binned by the number of observations per platform (top histograms in each panel), and by intensity category (bottom-right histogram). The 2D matrix plots are normalized by the respective overall maxima of populations for each observation type. TS, H, and HM stand for tropical storm, hurricane, and major hurricane, respectively.

have become more random in nature, but still suffer from a slight bias of underestimation.

Zonal and meridional wind observations are assimilated separately in HEDAS. For zonal wind, there appears to be a slight easterly (positive) bias (57% likelihood) for the first-cycle prior innovations (combined flight-level and dropsonde platforms). However, since the zonal wind observation is a vector, this result cannot be directly linked to intensity. (In other words, strong intensity is equally likely to result in positive, eastward, and negative, westward, zonal winds.) This bias is mostly removed in the final analysis: the likelihood of encountering

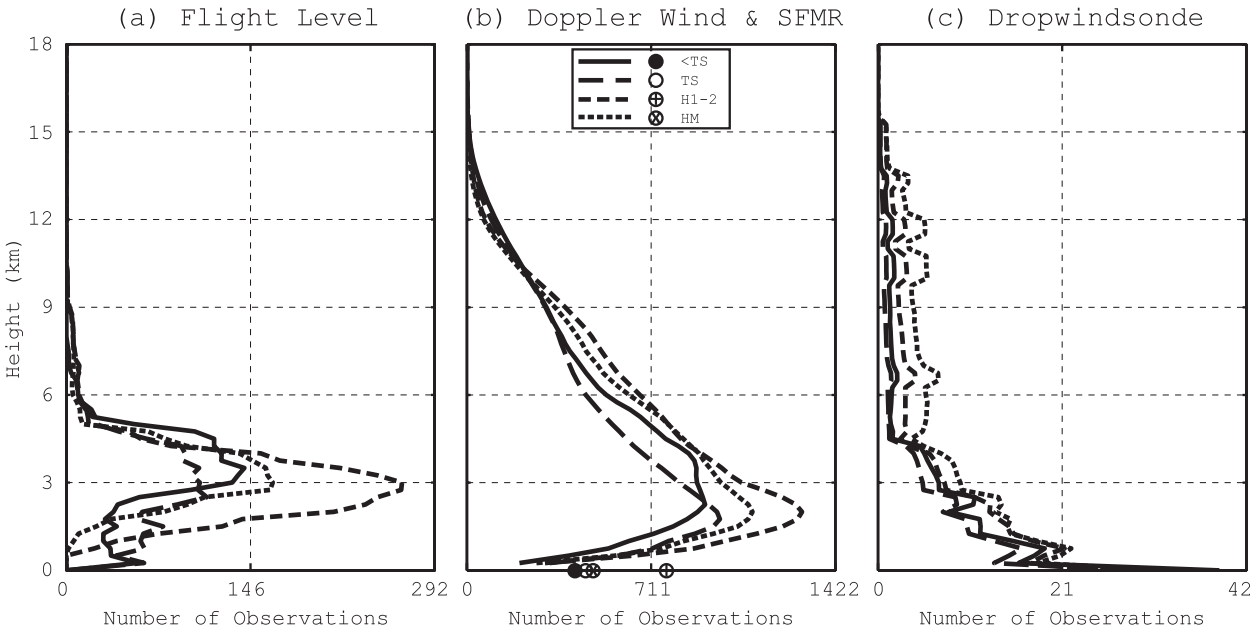

FIG. 4. Vertical distribution of number of observations for (a) flight-level, (b) Doppler wind and SFMR (circles for 0-m observations), and (c) dropsonde platforms. In each panel, distributions are categorized by storm intensity. Cases not in the best-track database (see Table 2) are assigned 20-kt intensity for plotting purposes.

underestimated easterly wind is down to 52%. The results for the meridional wind are similar due to the cyclonic nature of the phenomenon.

Finally, for temperature observations (combined flight-level and dropsonde platforms), a slight positive bias in the first-cycle prior innovations (55% likelihood) suggests that the warm core is underpredicted; that is, a weaker-than-observed background vortex is available to the EnKF in the first assimilation cycle. In the final analysis, the magnitude of this apparent bias is slightly reduced but the bias becomes negative.

2) VERTICAL VARIATION OF OBSERVATION INNOVATIONS

Figures 6–8 show the vertical variations of innovation statistics for Doppler wind speed, SFMR, temperature, and zonal wind speed as computed before the first assimilation cycle (i.e., first-cycle prior) and after the last assimilation cycle (i.e., last-cycle posterior). The 95% confidence intervals are also provided so that the statistical significance of the differences can be inferred. In computing the statistics, all observations of a specific type are aggregated from all cases that are within the intensity category of interest. Therefore, these statistics reflect the average behavior of the EnKF as applied to various intensity categories.

Figure 6 indicates that mean innovations of Doppler wind speed are generally small, even at the first cycle. Some biases of up to 1 m s^{-1} are apparent at mid- and upper levels of the troposphere (~4–14 km), which is reduced in the analyses at the last cycle. Much more prominent errors are observed in root-mean-square (RMS) innovations at the first cycle, which become gradually greater for storms of greater intensity (up to ~10–12 m s^{-1} for major hurricane cases). Clearly, there appears to be a strong correlation between storm intensity and the RMS departure of the background model field from Doppler wind speed observations. This is a result of weaker-than-observed storm intensity in the background wind field even after a spinup of 3–4 h from the global GFS/EnKF analysis ensemble. Meanwhile, the RMS innovations are reduced to the level of the observation error (2 m s^{-1}) in the final analysis, which is an indication of the effectiveness of the EnKF in obtaining a good fit to the inner-core observations in observation space.

Innovation statistics of SFMR observations depict a somewhat different picture than those of Doppler wind speed (Fig. 6, square symbols in each panel), as mean background–observation departures are somewhat greater in magnitude. Similar to Doppler wind speed, RMS departures still increase with increasing storm intensity and approach about 10 m s^{-1} for major hurricane cases. In the final analysis, mean innovations are reduced to below 1 m s^{-1} for all intensity categories and RMS innovations are reduced to approximately

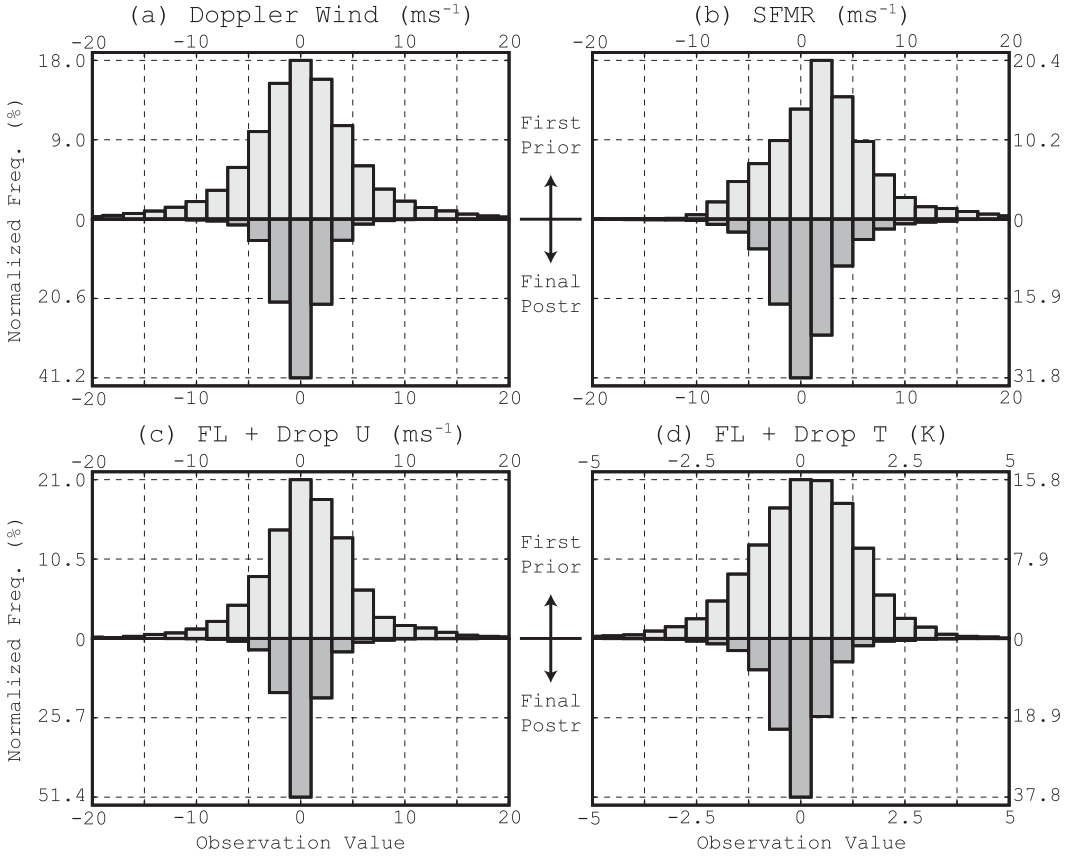

FIG. 5. Normalized frequency distribution (%) of innovations for (a) Doppler wind speed, (b) SFMR, (c) flight-level and dropsonde zonal wind speed U, and (d) flight-level and dropsonde temperature observations T. In each panel, prior distributions at the first assimilation cycle are shown in the top histogram, while posterior distributions at the final assimilation cycle are shown in the bottom histogram. Statistics are accumulated over all of the cases processed.

4 m s^{-1}, which is roughly the mean observation error for the SFMR platform.[2]

Although somewhat noisy in nature, innovation statistics for the zonal wind speed (as measured by aircraft at flight level or by dropsondes) are similar to the SFMR observations (Fig. 7). Innovations in the final analyses are distinctly smaller than those in the first-cycle background at most vertical levels. However, unlike the Doppler wind speed observations, RMS errors remain greater than the 2 m s^{-1} observation error in the lowest 1 km of the troposphere. While this hints at generally greater wind speed errors near the surface, it is also possible that the greatest impact on the near-surface innovations is from SFMR observations, which would partially explain the analysis innovations remaining near 4 m s^{-1}.

For temperature (measured by aircraft at flight level or by dropsondes), there appear to be negative biases of about 1–2 K below ~1 km and generally positive biases of about 1 K above ~1 km in the first-cycle background innovations (Fig. 8). The midlevel positive biases are especially noteworthy for weaker cases but are eliminated by data assimilation in the final analyses. Meanwhile, although reduced, negative low-level biases remain in the final analyses for the stronger cases. A similar trend is observed in the RMS innovations. For cases of tropical storm intensity or weaker, the RMS innovations in the final analyses are reduced back to approximately 0.5 K, comparable in magnitude to the observation error. However, stronger TC cases continue to exhibit RMS departures in the final analyses of 1–2-K magnitude, especially at lower levels.

Finally, the spread ratio (Fig. 6–8, right columns) measures the sufficiency of the ensemble spread in

[2] For SFMR observations, a variable observation variance is used that is rain-rate dependent. The dependence of the SFMR surface wind speed retrieval on the rain rate is explained in Uhlhorn et al. (2007). The mean SFMR observation error for all cases was ~5 m s^{-1}.

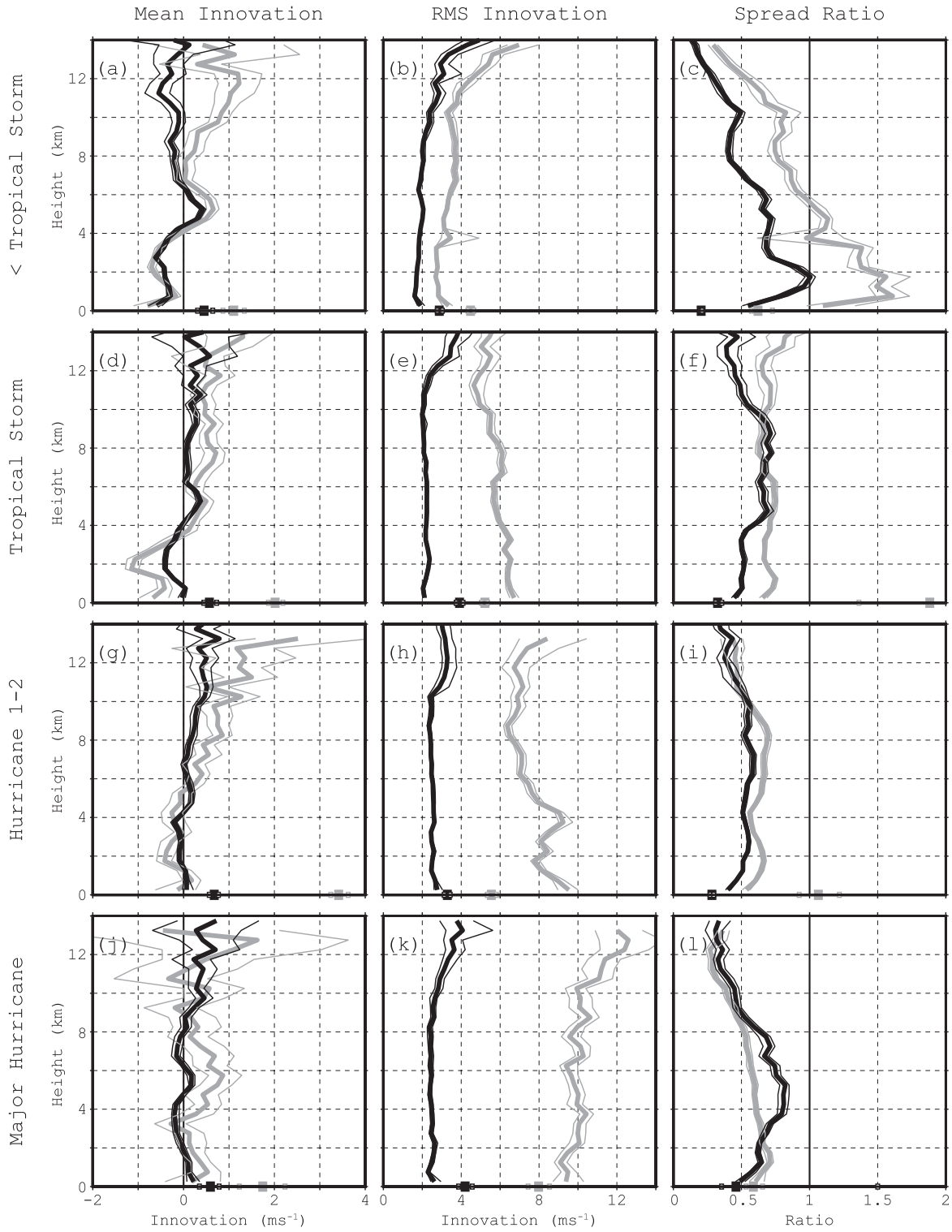

FIG. 6. Observation-space innovation statistics for Doppler wind speed and SFMR (squares for 0-m innovations) first-cycle prior (thick gray lines) and final-cycle posterior (thick black lines) distributions. Shown are (left) mean innovations, (middle) RMS innovations, and (right) spread sufficiency ratio. For spread ratio, the black lines represent final-cycle prior distributions. The statistics are aggregated for (a)–(c) weaker than tropical storm, (d)–(f) tropical storms, (g)–(i) category-1 and -2 hurricanes, and (j)–(l) major hurricanes. Thin lines represent 95% confidence intervals.

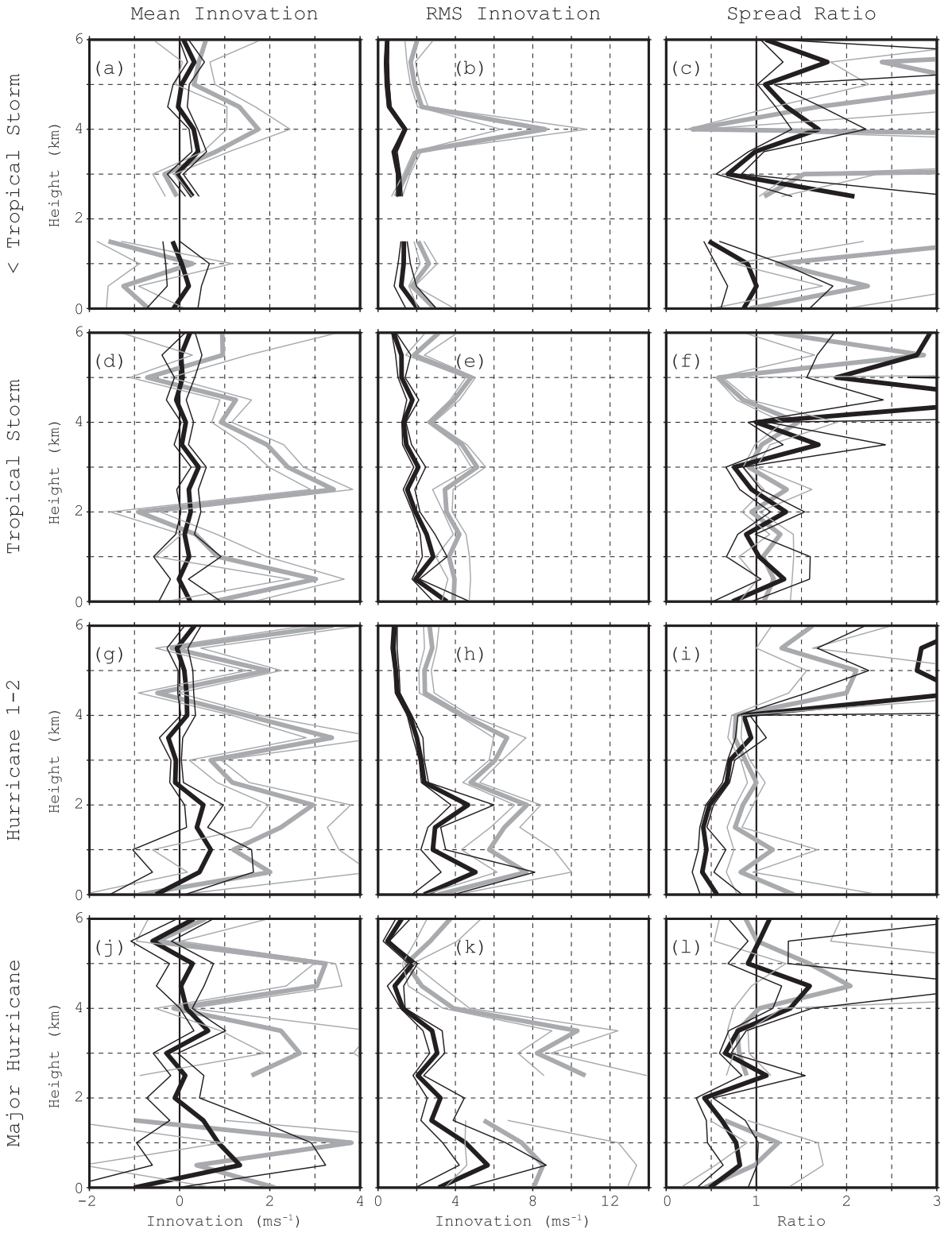

FIG. 7. As in Fig. 6, but for zonal wind speed observations. Gaps are due to the limited number of observations, in which situation statistics are not computed.

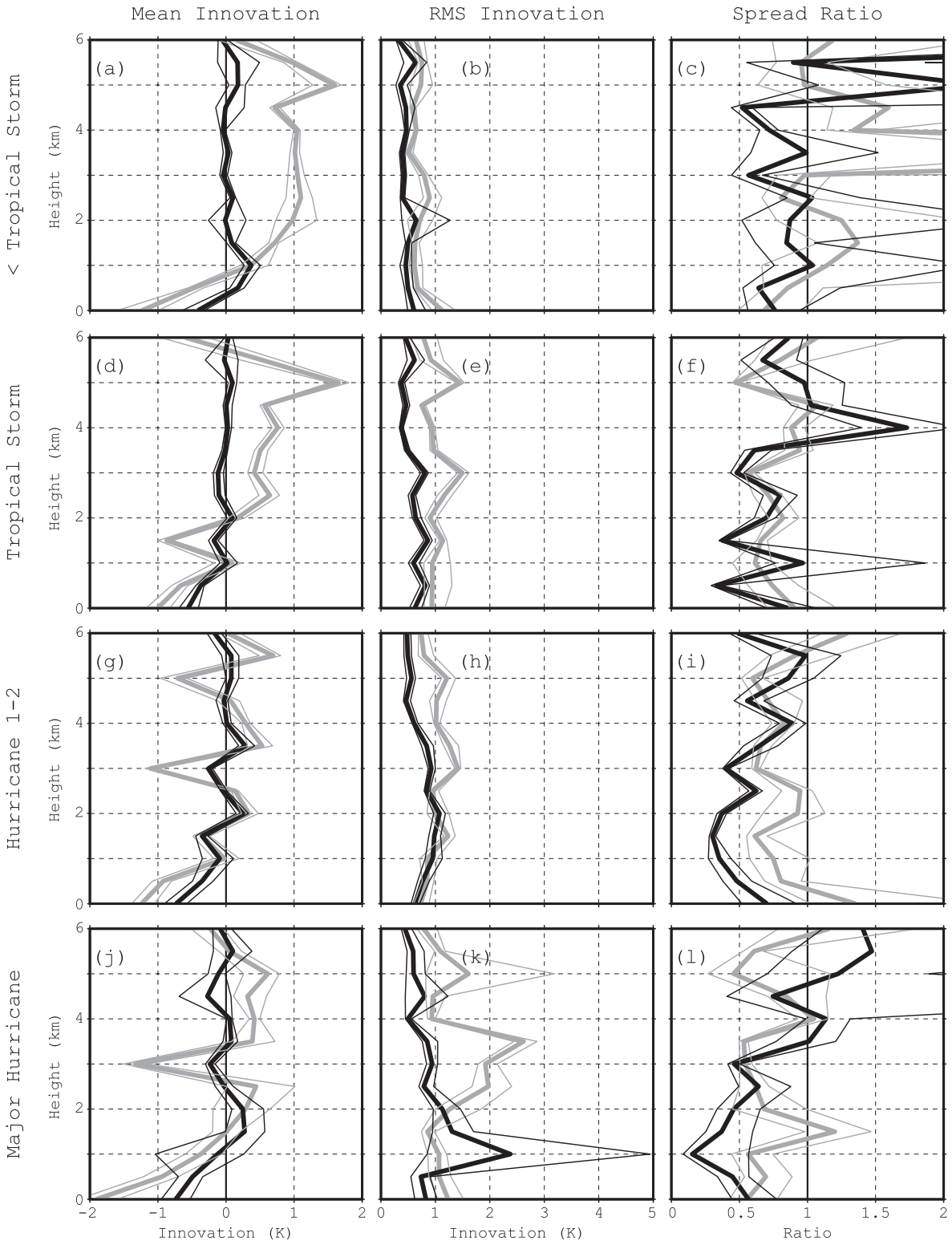

FIG. 8. As in Fig. 6, but for temperature observations.

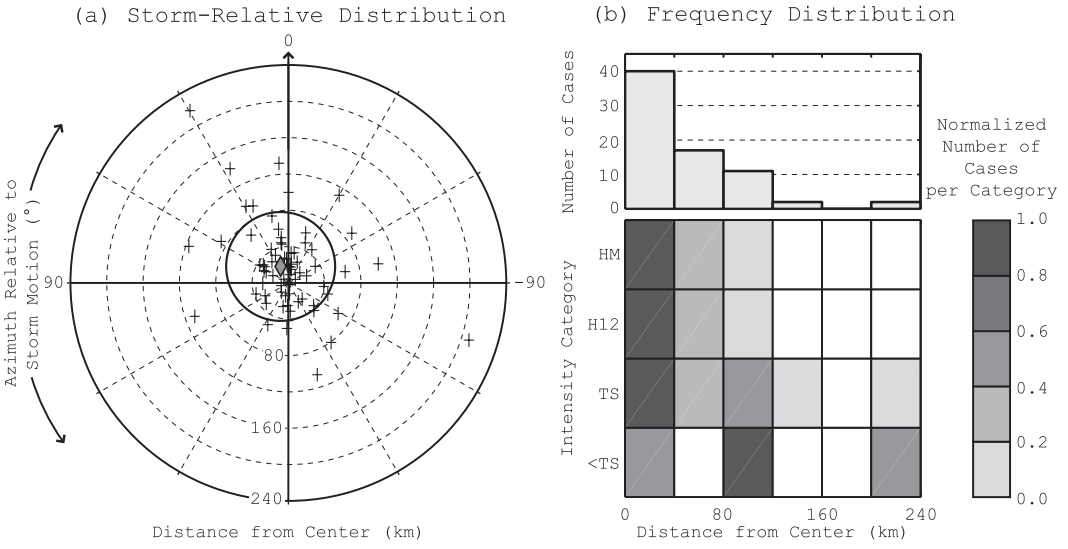

FIG. 9. Position error in the final analysis as compared to the best track. (a) Analysis storm centers (plus markers) relative to the best track. Azimuth is measured relative to observed storm motion where 0° represents the direction of storm motion. Radial distance is measured from the best-track storm center. The centroid location of all cases is shown with the diamond marker. The standard deviation of position errors is indicated with the circle around the centroid location. (b) Number of cases as a function of the analysis–observed radial distance of storm centers and intensity category (2D matrix plot, gray scale on right), and cumulative as a function of the analysis–observed radial distance of storm centers (histogram). The 2D matrix plot population bin values are normalized by the respective maxima of populations for each intensity category.

comparison to random forecast error and observation error. As explained in Aksoy et al. (2009), the optimal ratio is 1, with smaller values reflecting insufficient ensemble spread. Similar to the A12 findings, the HEDAS ensemble appears to be somewhat deficient in spread, especially for Doppler wind speed observations. However, one important difference between the HEDAS implementation here and that in A12 is that, unlike A12, no covariance inflation is applied for the real-data cases presented in this work. Therefore, it is only logical to infer that what is lost in ensemble variability throughout cycling from not applying any covariance inflation is gained from using the GFS/EnKF analysis ensemble perturbations at the cold start rather than the operational GFS ensemble perturbations that are used in A12. Nevertheless, it is noted that spread sufficiency values generally vary in the 0.5–1.0 range for all observed variables.

b. Position and intensity compared to best track

In this study, the HEDAS final analyses (to be referred to as analyses hereafter) are verified by comparison with the best-track analyses at the nearest synoptic time. As the time of the final analyses is somewhat arbitrary and depends on when the Doppler radar measurements end, up to a 3-h difference is possible between the verification and final analysis times. This introduces a slight overestimation of analysis errors, especially for storm position, which is temporally progressive in nature.

The distribution of analysis TC center locations [defined here as the locations of minimum sea level pressure (MSLP)] compared to those from the best track is shown in Fig. 9a. For a relevant storm-relative comparison, the analysis–best-track displacements are shown with respect to the storm motion direction (calculated from the best track centered at the nearest synoptic time). It is inferred from the centroid of displacements that analyses exhibit a slight left and forward bias relative to the best-track storm position. When the distribution of the distance between analysis and best-track storm centers is analyzed (Fig. 9b), most of the cases are found to exhibit 40 km or smaller position errors. (When best-track position is interpolated into the final analysis time, a mean position error of 38 km is obtained; not shown.) Moreover, the cases with the greatest position errors (~100 km or greater) are of tropical storm intensity or weaker. This is likely the result of a combination of two possible scenarios. 1) The HEDAS system is better capable of analyzing storm position in stronger storms. This would mainly result from better-defined radial gradients that lead to stronger correlations between wind observations and position. 2) The general procedure of center finding itself is easier (and therefore more accurate) in stronger storms, mainly because of stronger

radial gradients and fewer local surface pressure minima in stronger storms.

In terms of intensity, the analyses compare well against the best track. Figure 10a shows that, for maximum 10-m wind speed, which is the standard measure of intensity, the coefficient of determination R^2 is 87% (i.e., the analyses explain 87% of the variance in the best track). There is also a 1.1 m s^{-1} negative bias in analysis intensity (i.e., underestimation of intensity), although it is not statistically significant at the 95% confidence level. An even better linear regression fit is achieved for MSLP (R^2 = 97%; Fig. 10b). However, this is now accompanied by a more distinct, positive bias of 3.7 hPa (underestimation of intensity) that is statistically significant. The analysis–best-track similarities in maximum intensity and MSLP are also reflected in the wind–pressure relationship (Fig. 10c). For the cases analyzed here, both HEDAS and the best track depict a linear relationship with R^2 = 82%–83%. The separation between the two regression lines reflects the positive bias in analysis MSLP, although it is not statistically significant at the 95% confidence level (not shown).

It should be noted that while the analysis position and intensity are compared here against those in the best track, the best-track estimates themselves also contain significant uncertainties. Torn and Snyder (2012) discuss that best-track position accuracy varies with storm intensity from ~40 n mi (74 km) for tropical depressions to ~15 n mi (28 km) for category-4 and -5 hurricanes. Although HEDAS position error demonstrates a similar dependence on intensity, generally greater values of average HEDAS–best-track deviations than the best-track uncertainty estimates by Torn and Snyder (2012) indicate smaller position accuracy in HEDAS than the best track. For intensity, Torn and Snyder (2012) point out a generally flat intensity uncertainty profile at 9–12 kt (~4.5–6 m s^{-1}) for storms of tropical storm or greater intensity. A generally similar trend is seen in Fig. 10a for both the HEDAS analyses and the best track, where the uncertainties in the intensity estimates against the linear regression line are computed as 9.3 and 9.8 m s^{-1}, respectively.

c. *Primary circulation compared to airborne Doppler radar observations*

The focus is now turned to the properties of the primary (i.e., azimuthal) circulation structure in the HEDAS analyses as compared against those obtained from airborne Doppler radar composite wind retrievals. Three-dimensional wind retrievals of NOAA P-3 tail Doppler radar data using a variational method (Gamache 1997; Gao et al. 1999; Reasor et al. 2009) are routinely performed aboard the aircraft after each center penetration.

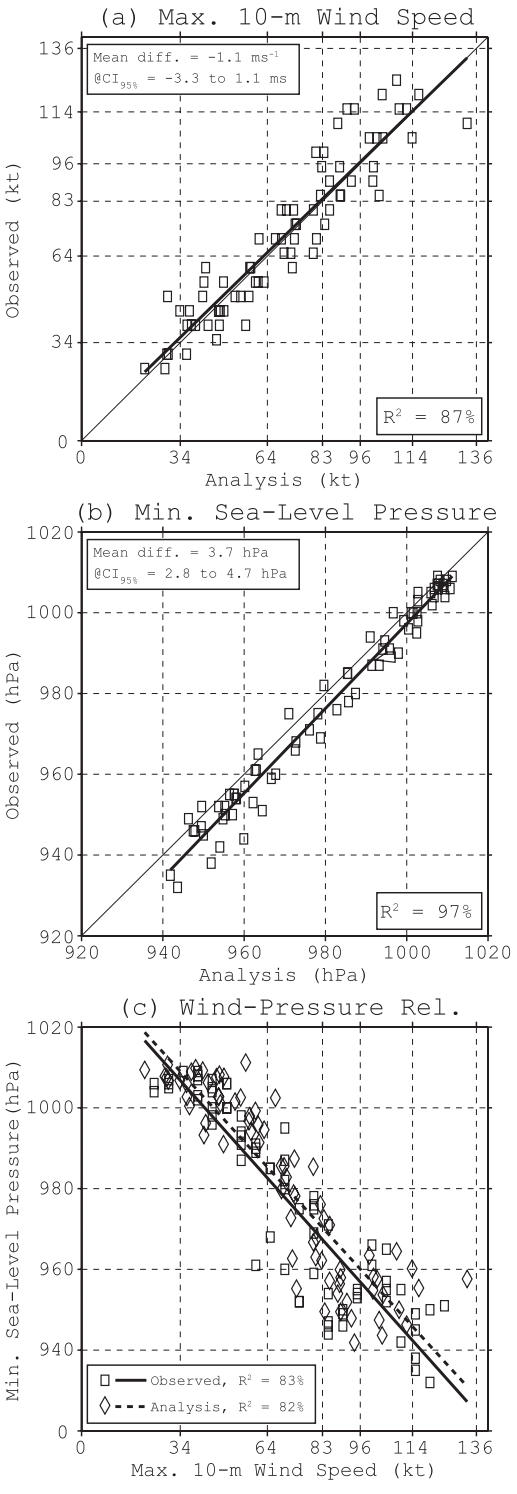

FIG. 10. Intensity error in the final analysis as compared to the best track. (a) Analysis vs observed scatter diagram of maximum 10-m wind speed (kt) for all cases in the best-track database. The solid black line represents the linear regression between the analysis and observations. The gridlines represent intensity category thresholds. (b) As in (a), but for MSLP. (c) Wind–pressure relationship in observed (square markers, solid linear regression line) and analysis (diamond markers, dashed linear regression line) data.

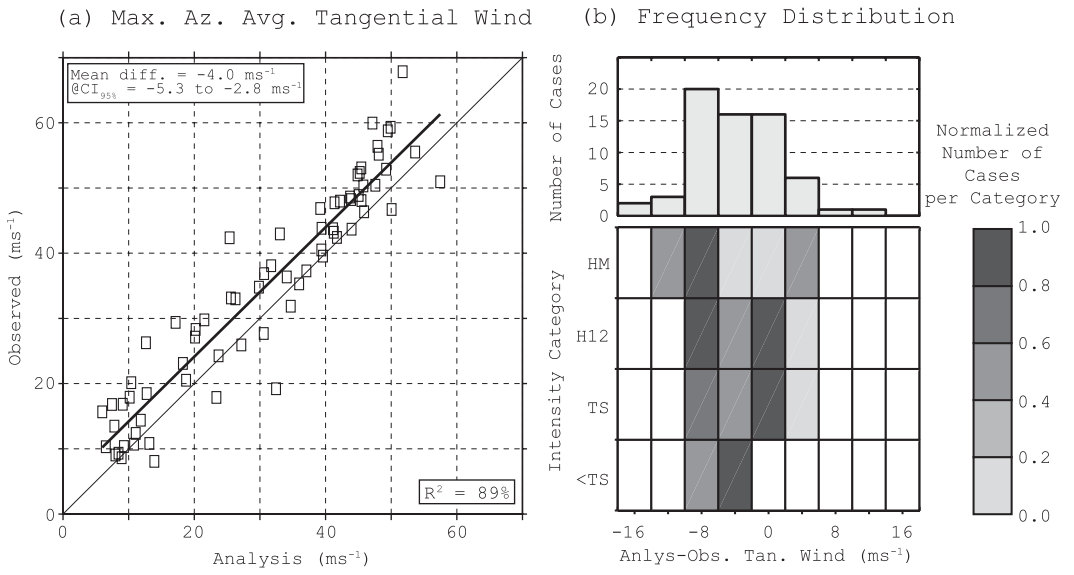

FIG. 11. Maximum azimuthally averaged tangential wind speed (m s^{-1}) as compared to radar observations. (a) Analysis vs observed scatter diagram for all cases in the best-track database. The solid black line represents the linear regression between the analysis and observations. (b) Number of cases as a function of analysis–observation difference and intensity category (2D matrix plot, gray scale on right), and cumulative as a function of analysis–observation difference (histogram). The 2D matrix plot population bin values are normalized by the respective maxima of populations for each intensity category.

No background vortex information is used in the retrievals produced in this way. After the completion of all such retrievals for a particular flight, they are composited within a storm-relative framework: for each grid point, the wind speed with highest magnitude is found, and the zonal and meridional components of the wind are computed. Therefore, they reflect the characteristics of the observed vortex structure as closely as possible at available data points. As Reasor et al. (2009) explain, the smallest uncertainties of the tangential and radial wind speed retrievals, in the range of 2–3 m s^{-1}, are found in the eyewall region and below 3-km altitude where most scatterrers are encountered.

It should be noted in advance that the quality of the computation of the structural characteristics of the primary circulation discussed here depends greatly on the accuracy of the storm position itself. As was mentioned in the previous section, the storm position accuracy varies with storm intensity and is greatest for the weakest tropical cyclones. Therefore, a general increase in the uncertainty of the structural characteristics with intensity is expected.

In terms of the maximum azimuthally averaged tangential wind speed (at any altitude), Fig. 11a shows a robust linear fit between analyses and radar observations with $R^2 = 89\%$. However, a statistically significant negative bias of 4.0 m s^{-1} is also apparent, indicating that, in an azimuthally averaged sense, HEDAS has a tendency to underestimate the strongest part of the primary circulation. It is also worthwhile to note that a more distinct bias is apparent in maximum azimuthally averaged tangential wind speed than in maximum 10-m wind speed. As was discussed in A12, this reflects the noisier nature of the maximum 10-m wind speed as a measure of overall TC intensity.

The dependence of the analysis–observed intensity difference is also analyzed as a function of observed intensity. Figure 11b shows the frequency distribution of the number of cases for each intensity category as a function of analysis–observation difference in maximum azimuthally averaged tangential wind speed. It is evident that the mode of the probability distribution shifts toward more negative values (more underestimation) with increasing intensity. Furthermore, the widening distributions with increasing intensity point to the greater degree of disagreement between HEDAS and radar observations on TC intensity. Possible causes of such systematic errors in the HEDAS analyses are investigated in Vukicevic et al. (2013).

Analyses are also compared against how well they represent the structure of the primary circulation. Being purely observational in nature, the gridded radar retrievals contain data gaps due to instrument limitations. In some cases, such gaps can lead to uncertainties in determining the structural properties of the vortex. Lorsolo and Aksoy (2012) provide a discussion on the impacts of

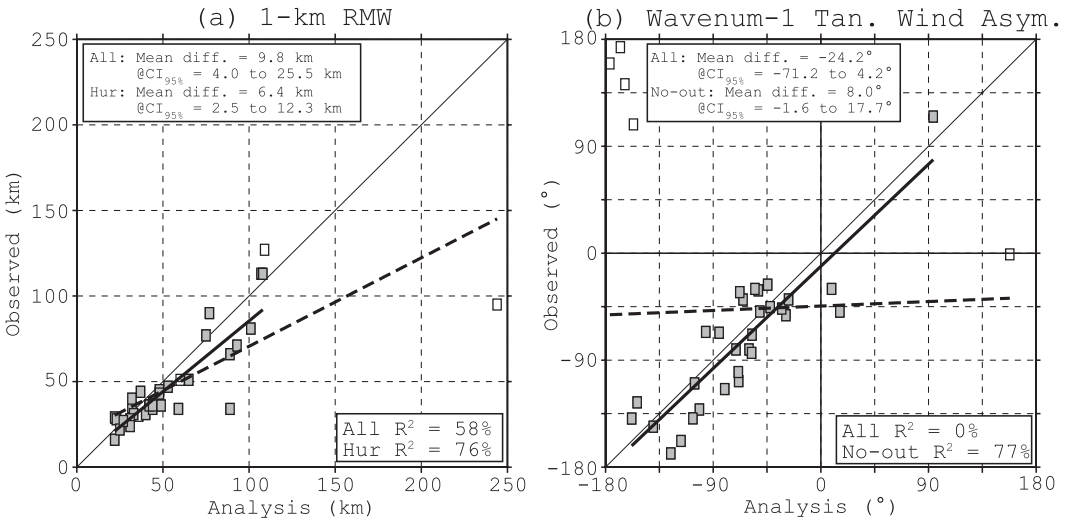

FIG. 12. (a) Scatter diagram of analysis vs radar-observed RMW (km) at 1-km altitude for all cases with maximum gap size of 150° or less (see text). Filled squares represent hurricanes. The thick dashed (solid) line represents the linear regression between analyses and observations for all (hurricane) cases. (b) As in (a), but for the azimuthal phase of wavenumber-1 asymmetry (° from storm motion) at 1-km altitude and one RMW radial distance. Filled squares represent cases that are deemed not to be outliers (see text). The thick dashed (solid) line represents the linear regression between the analysis and observations for all (nonoutlier) cases.

gap size on wavenumber analysis of azimuthally distributed data. Here, for a fair comparison of model and observed radar data, cases with relatively large maximum azimuthal gap size in the radar data are not considered in the investigation of the primary circulation structure (Figs. 12 and 13). Following the recommendations in Lorsolo and Aksoy (2012) for a signal that is either wavenumber-0 or wavenumber-1 dominant (as is generally the case for the tangential wind speed) and at least with two gaps present in the data, a maximum gap size of 150° at 1-km altitude and a radial distance of one radius of maximum azimuthally averaged tangential wind (RMW) is allowed for the retrieval of these wavenumbers. This threshold resulted in 47 of the 83 cases (57%) being disregarded for the rest of this section. All comparisons in Figs. 12 and 13 are carried out at 1-km altitude, where the tangential wind speed is typically maximized (J. Zhang et al. 2011, their Fig. 10; Rogers et al. 2012, their Fig. 7).

In Fig. 12a, the RMW is investigated. When all cases are considered, $R^2 = 58\%$ (dashed line). However, a much better agreement is achieved ($R^2 = 76\%$) when hurricane-only cases are considered (solid line). This shows that the ability of HEDAS to estimate the size of the inner-core vortex, as measured by the RMW, is clearly a function of observed intensity.

The structure of the tangential wind is further investigated through the azimuthal phase of the wavenumber-1 asymmetry at the RMW (Fig. 12b). No linear statistical relationship is discernible between analyses and radar observations ($R^2 = 0\%$, dashed line). Unlike RMW, wavenumber-1 asymmetry correlation between HEDAS and observations does not appear to respond to intensity either: for the hurricane-only cases, the variance explained remains very low at 2% (not shown). However, there exist some noticeable outlier pairs in the wavenumber-1 asymmetry scatter diagram (Fig. 12b). When a threshold of 90° is applied for the maximum absolute difference to be allowed between analyses and radar observations (roughly limiting the azimuthal phase to the same quadrant), five cases (14% of all cases) are deemed outliers. Without these outlier cases, a much better degree of statistical agreement between the analyses and observations is achieved ($R^2 = 77\%$, solid line). The outlier cases do not reveal distinct common features: no preference for characteristics such as intensity, geographical location, and RMW are found (not shown). It is inferred that the occurrence of these outliers is mostly random in nature.

The primary circulation structure is also examined in terms of the azimuthal wavenumber-0 and -1 components (Fig. 13). Wavenumber amplitudes are obtained at the RMW while the variance explained by the wavenumber components is computed for a radial range of 0–300 km. The magnitude of the wavenumber-0 amplitude (Fig. 13a) reveals a very good linear fit at $R^2 = 89\%$. A statistically significant negative bias of 2.4 m s^{-1} is also apparent, indicating that analyses tend to slightly underestimate the intensity of the mean azimuthally averaged tangential wind speed. These results are consistent

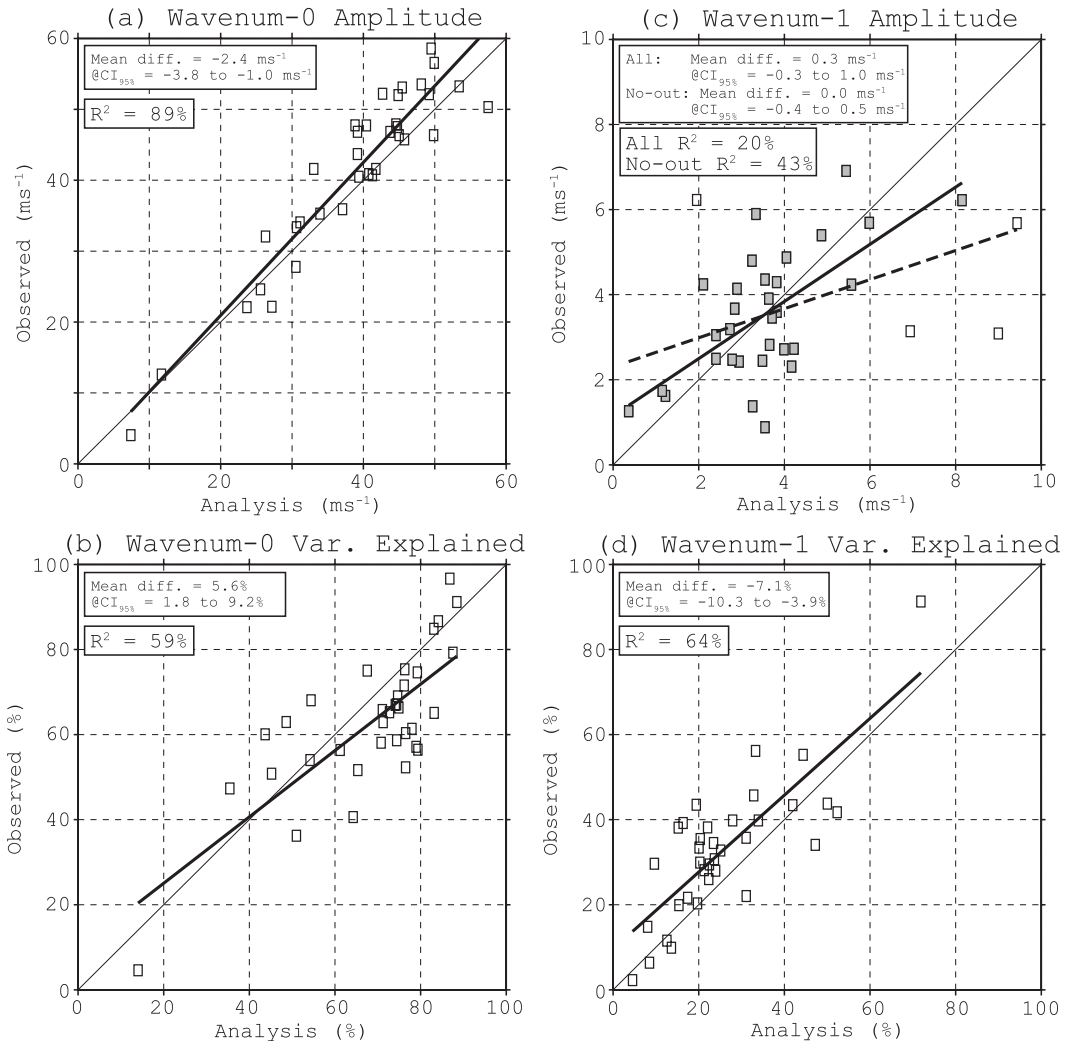

FIG. 13. Analysis vs radar-observed scatter diagram of azimuthal wavenumber-0 and -1 characteristics of azimuthally averaged tangential wind speed at 1-km altitude for all cases with maximum gap sizes of 150° or less (see text) in wavenumber-0 (a) amplitude (m s^{-1}) and (b) variance explained (%). (c),(d) As in (a),(b), but for wavenumber-1. The solid lines represent the linear regressions between the analysis and observations for each parameter. In (c), filled squares represent cases that are deemed not to be outliers (see text). The dashed line represents the linear regression between the analysis and observations for nonoutlier cases. In (a) and (c), computations are at 1 RMW while in (b) and (d), a radial range of 0–300 km is used.

with those for the maximum azimuthally averaged tangential wind speed (Fig. 11a). That a greater negative bias of 4 m s^{-1} exists for the *maximum* of this parameter is an indication that HEDAS could have more difficulty in capturing the extrema of the primary circulation than its average properties.

The variance explained by the wavenumber-0 component of the tangential wind speed (i.e., how well the wavenumber-0 component is correlated with the full field) is also compared between analyses and radar observations (Fig. 13b). A relatively good statistical fit of R^2 = 59% is obtained. On average, HEDAS appears to overestimate the variance explained by the wavenumber-0 component by 5.6%. The overestimation of the wavenumber-0 component in HEDAS is accompanied by underestimation of the variance explained by the wavenumber-1 component by a comparable magnitude (Fig. 13d). It appears that the signal that HEDAS tends to allocate to the wavenumber-0 component occurs at the expense of the signal in the wavenumber-1 component. In other words, there is a small but systematic tendency (5% of the total variance) in HEDAS to axisymmetrize the primary circulation.

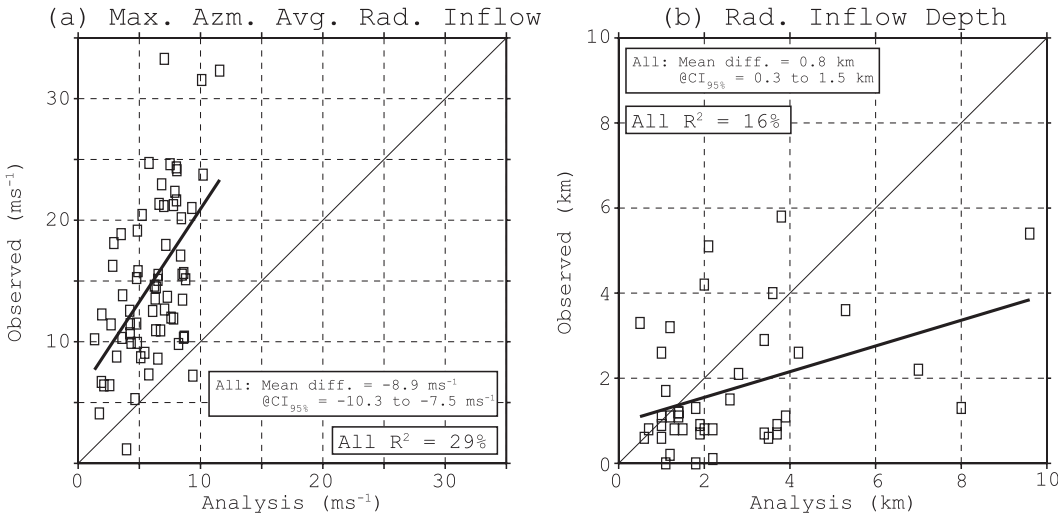

FIG. 14. (a) Scatter diagram of analysis vs radar-observed maximum azimuthally averaged radial inflow (m s^{-1}) for all cases in the best-track database. The black solid line represents the linear regression between the analyses and observations for all cases. (b) As in (a), but for the depth of the radial inflow (km).

Finally, analyses are found to be the least effective in representing the amplitude of the wavenumber-1 component of the tangential wind speed. For this parameter, although eliminating outlier cases (absolute difference between analyses and observations of 3 m s^{-1} or greater) improves somewhat the linear fit between analyses and radar observations, at $R^2 = 43\%$, the linear relationship is still quite weak. As the variance explained by the wavenumber-1 component is found to exhibit a much better linear fit, it is concluded that the assimilation of radar observations in HEDAS has a stronger influence on the spatial pattern of the wavenumber-1 asymmetry than its amplitude. Furthermore, when outlier cases are not considered, the average wavenumber-1 amplitude magnitude of 3.5 m s^{-1} for analyses (same for radar observations) is comparable to the standard error of the wavenumber-0 amplitude of 1.9 m s^{-1} (2.1 m s^{-1} for radar observations). It is therefore conjectured that the wavenumber-1 amplitude exhibits a low signal-to-noise ratio compared to the variability of its wavenumber-0 counterpart.

d. Secondary circulation compared to radar observations

Here, only the low-level radial inflow component of the secondary circulation is investigated, as there are very few observations in the upper-level outflow region to carry out statistically sound comparisons and the vertical motion is not analyzed in HEDAS (see A12 for details).

Estimating the secondary circulation with HEDAS is found to be more challenging than the primary circulation. This is partially as a result of the uncertainty in the storm center estimation itself, as this uncertainty significantly impacts the robustness of the computation of the radial wind especially near the RMW. Nguyen and Molinari (2012) demonstrate this effect nicely in their Fig. 9.

Figure 14 compares the magnitude (Fig. 14a) and the depth (Fig. 14b) of the maximum azimuthally averaged radial inflow for analyses and radar observations. For all cases considered, analyses underestimate the magnitude of the radial inflow by 8.9 m s^{-1} and overestimate its depth by 0.8 km. Both estimates are statistically significant at the 95% confidence level. Bao et al. (2012) documented the systematically positive inflow depth bias in the experimental HWRF model (cf. their Fig. 4a, which represents the experimental HWRF configuration, to their Fig. 4b) that arises from the choice of the physical parameterizations, while Gopalakrishnan et al. (2013) demonstrate the impact of eddy diffusivity on the vertical structure of the boundary layer in the same model. It is believed that the findings here are consistent with those findings. For a given magnitude of mass convergence in the boundary layer, this systematically deeper inflow in HEDAS is then translated to weaker mass flux in the boundary layer.

e. Composite radial profiles

In Figs. 15–18, the azimuthally averaged kinematic and thermodynamic structure in the analyses is investigated in a composite sense and compared to corresponding composites obtained from P-3 observations at flight level and Real-time Hurricane Wind Analysis System (H*Wind) analyses at the surface (Powell et al. 1998).

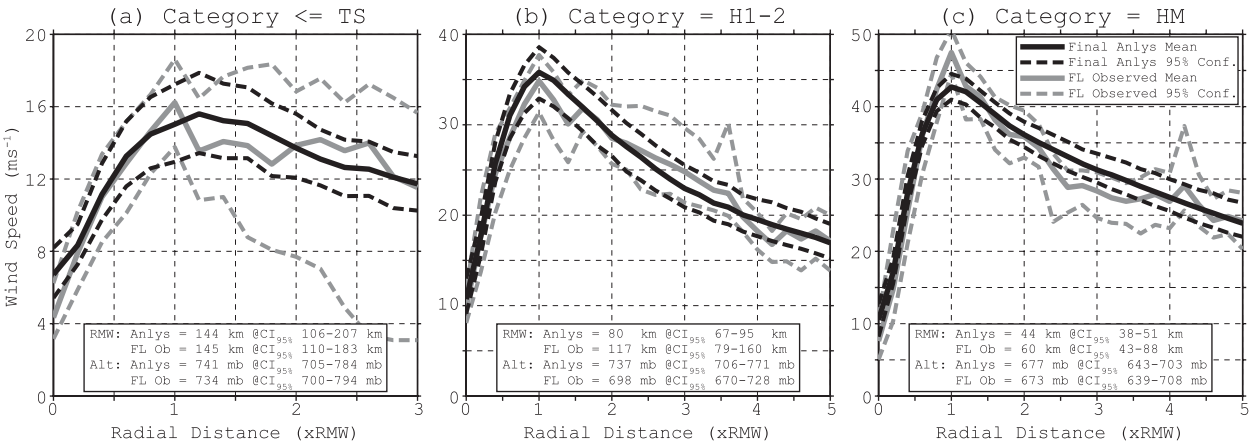

FIG. 15. Composite radial profiles of azimuthally averaged horizontal wind speed (m s^{-1}) at P-3 flight level for HEDAS final analyses (black solid) and in situ aircraft observations (gray solid) for (a) all cases that had tropical storm or weaker intensity in the best-track database, (b) all cases with category-1 and -2 hurricane intensities, and (c) major hurricanes. The 95% confidence intervals are shown as dashed lines. Average RMW and flight altitude (Alt) in the analyses as well as in the observations are shown in the box inserts, along with corresponding 95% confidence intervals.

Composite radial profiles of wind speed, temperature, and specific humidity are computed as follows. First, analysis fields (both HEDAS and H*Wind) are normalized in the radial direction by RMW and then transformed to polar coordinates using the flight-level vorticity centroid as the center. These fields are sampled at 60° azimuthal resolution at the reported approximate pressure altitude of the flights to mimic the usual sampling strategy during P-3 flights. Flight-level radial profiles are similarly normalized by the RMW. RMW is computed separately for observations and HEDAS analyses at flight level and for H*Wind analyses at the surface.

It should be noted that the average RMW and altitude values reported for flight level and the surface in the inserts of Figs. 15 and 18, respectively, are generally consistent between observations and analyses. Overall, sufficient agreement exists between analyses and observations for RMW and flight altitude so that their comparison for normalized radial structures in Figs. 15–18 is meaningful. Some inconsistencies exist for RMW, especially for the weaker cases, which is in accordance with the previous findings in Fig. 12a.

At flight level, the mean structures of the horizontal wind speed are very similar for all cases up to hurricane category 2 (Figs. 15a,b) and statistically are not distinguishable at the 95% confidence level. For stronger hurricanes (Fig. 15c), analyses underestimate peak wind speed by approximately 5 m s^{-1} but overestimate wind speed farther out at about 3 RMW. It is also interesting to note that, except for weaker cases outside the RMW, the statistical uncertainty of wind speed at flight level is comparable between observations and analyses.

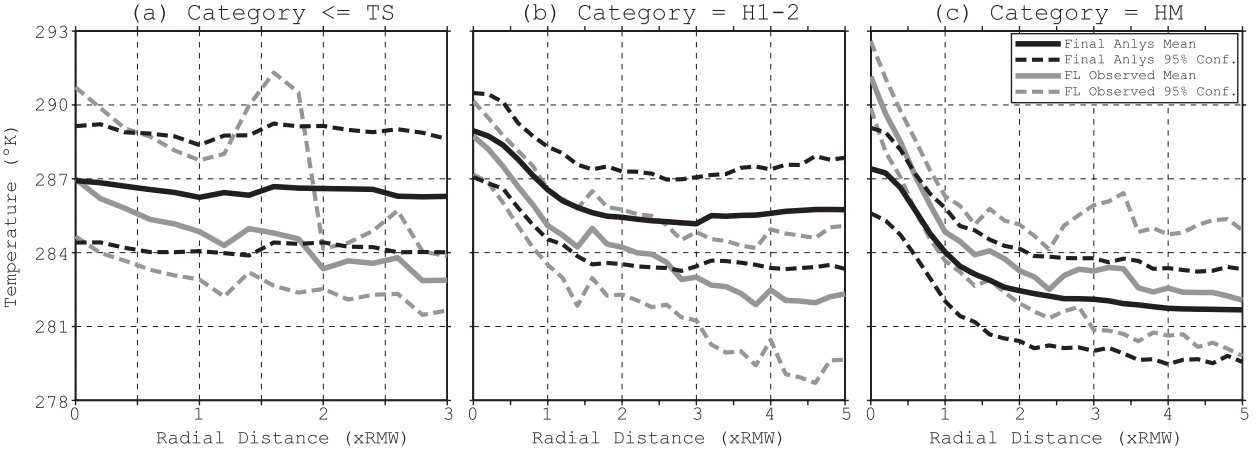

FIG. 16. As in Fig. 15, but for temperature (K).

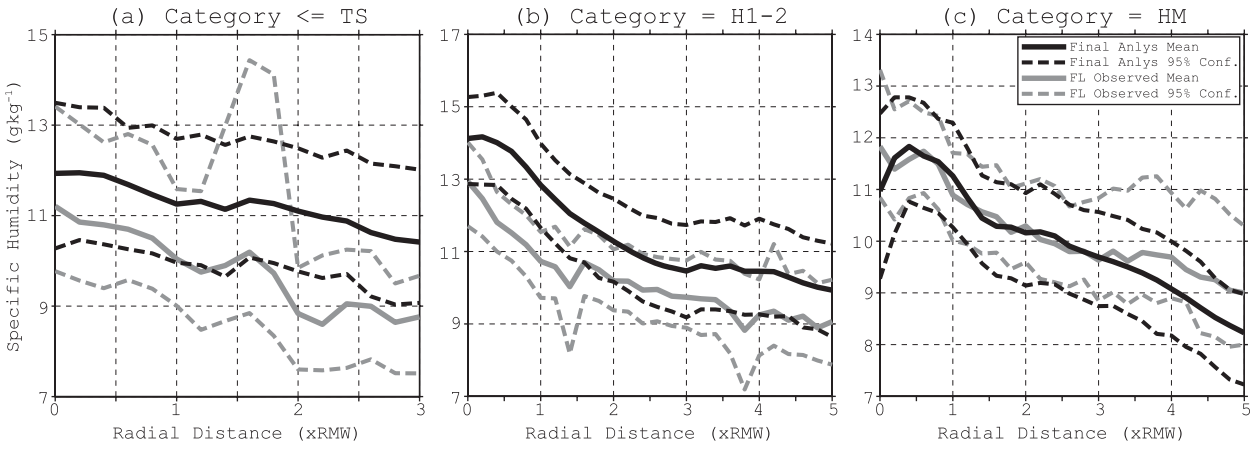

FIG. 17. As in Fig. 15, but for specific humidity (g kg^{-1}).

The thermodynamic structure at flight level reveals a more biased picture for the analyses, especially for the weaker cases. Figures 16a,b (weaker cases) show that while analyses capture the warm-core temperature at the storm center well, a warm bias of approximately 3 K exists outside the core. For major hurricanes (Fig. 16c), the picture reverses and analyses underestimate the warm-core temperature by approximately 4 K but capture well the temperature structure outside the core. A similar scenario is also observed for specific humidity, for which an overall overestimation of approximately 1 g kg^{-1} by HEDAS is evident for the weaker cases (Figs. 17a,b). Meanwhile, the structure is captured relatively well for major hurricanes (Fig. 17c).

The systematic overestimation of temperature and specific humidity in analyses is likely caused by a combination of multiple factors. The update of the thermodynamic fields in HEDAS relies heavily on the quality of information obtained indirectly by the correlations with Doppler wind observations. Although flight-level and dropsonde temperature observations are assimilated, they are sparse and their impact is limited to the cyclone core, likely because of the better spatial resolution in observations achieved in the core. Outside the core, limited impact from temperature observations reveals more clearly the underlying systematic tendencies in HEDAS (Vukicevic et al. 2013). The systematically deeper boundary layer that was discussed in section 5d also increases the probability of a warm and moist bias at any given altitude in the boundary layer. However, it is interesting that these biases become less pronounced outside the core for major hurricanes despite persistent boundary layer depth biases. Meanwhile, the distinct underestimation of the warm core for these cases appears to be consistent with the underestimation of kinematic intensity.

Finally, at the surface, analyses are in good general agreement with H*Wind analyses (Fig. 18). Underestimation

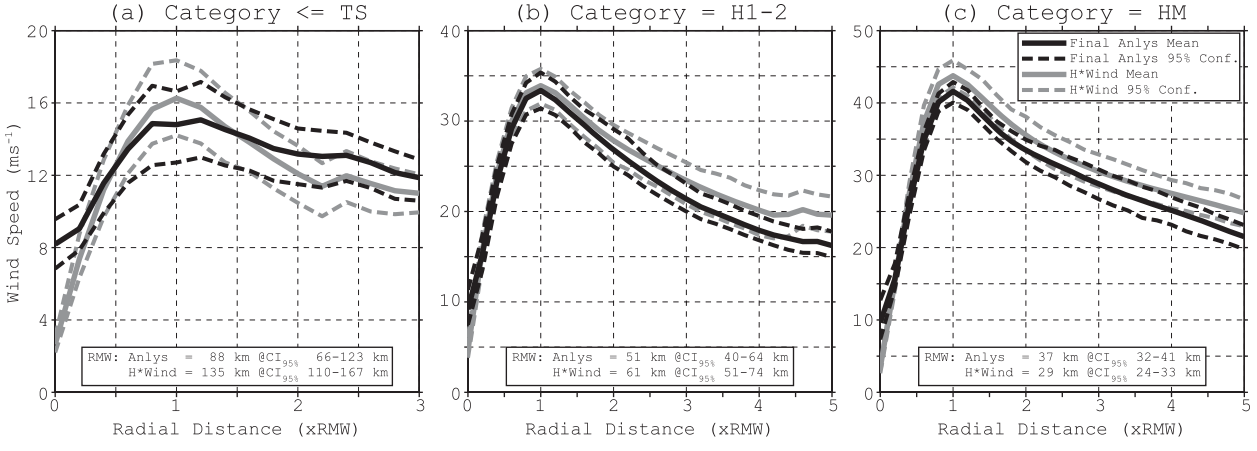

FIG. 18. As in Fig. 15, but for surface (10 m) wind speed (m s^{-1}). Here, HEDAS analyses are compared to H*Wind analyses.

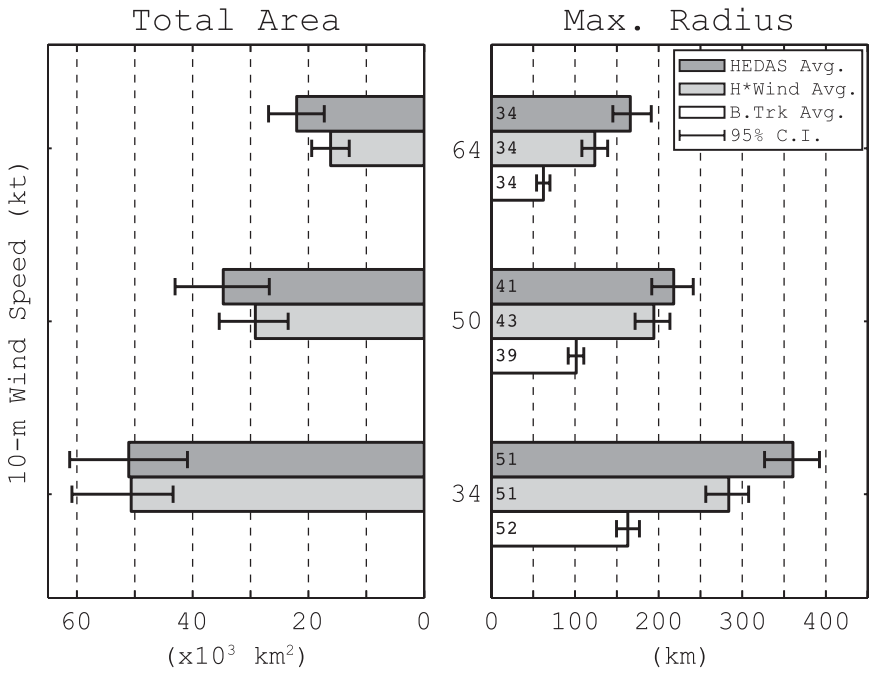

FIG. 19. Horizontal box chart that shows case-averaged storm size in HEDAS, H*Wind, and the best-track database for 34-, 50-, and 64-kt wind speeds at 10 m. Storm size is measured by (left) total area ($\times 10^3$ km^2) and (right) maximum radius (km). Number of cases considered in computing the averages are shown on the left of the maximum radius boxes. Statistical significance is indicated by the 95% confidence intervals.

of peak wind speed by approximately 2 m s^{-1} for major hurricanes in HEDAS analyses is consistent with previous findings.

f. Storm size

Figure 19 shows a comparison of storm size among HEDAS, H*Wind, and the best-track database for 34-, 50-, and 64-kt wind speeds at 10 m. For a more comprehensive understanding, storm size is computed as both the maximum radius of the respective wind speed from storm center in any quadrant and the total area spanned by the respective wind speed.

In terms of both metrics considered, largest (smallest) vortices are found in HEDAS (best track). Moyer et al. (2007) provide a thorough comparison of gale radius statistics between H*Wind analyses and best-track estimates and show that objective estimates of tropical storm force wind radii in H*Wind analyses were consistently larger than the best track. This finding appears to be consistent with Fig. 19. However, Moyer et al. (2007) also mention that when storm size was subjectively compared against observations in individual cases, it was found to be consistently underestimated in objective analyses from H*Wind because of the smoothing associated with the objective analysis procedure. This last finding suggests that the relatively large storm size seen in the HEDAS analyses may very well be realistic, although a more detailed investigation would be needed for a quantitative verification.

6. Summary and discussion

The impact of aircraft inner-core TC observations on analyses of high-resolution vortex structure is investigated through the use of NOAA/AOML/HRD's EnKF-based data assimilation system, HEDAS. Eighty-three cases (20 individual TCs) from years 2008 to 2011 are considered. With the exception of Hilary (2011) in the eastern North Pacific basin, all cases are in the Atlantic basin. Observed intensity ranges from tropical depression to category-4 hurricane, with most cases falling into the tropical storm category.

HEDAS assimilates available observations from NOAA P-3, G-IV, and/or AFRC C-130 flights at 1-h intervals. Among the 83 cases, 4 h of cycling is the most frequently encountered length of assimilation period. Assimilated observation types include aircraft Doppler radar wind superobs, temperature, and zonal and meridional wind from aircraft flight-level measurements and dropsondes, and SFMR 10-m wind speed retrievals. Among these

observation types, Doppler wind superobs are about one order of magnitude greater in number than the other types of aircraft data.

Observation-space diagnostics reveal some underdispersiveness in the HEDAS ensemble, especially for Doppler superobs. Although this finding is similar to that of A12, one important difference between the HEDAS implementation here and that in A12 is that, unlike A12, no covariance inflation is applied. Therefore, it is inferred that what is lost in the ensemble variability from not applying any covariance inflation is gained from using the GFS/EnKF analysis ensemble perturbations at the cold start rather than the operational GFS ensemble perturbations that were used in A12. The remaining ensemble spread deficiency is likely caused by model error that is not accounted for in the forecast ensemble. Covariance inflation techniques may have limited success in ameliorating this shortcoming. This was also demonstrated in A12, where the application of various covariance inflation techniques reduced but did not totally eliminate the spread deficiency.

When compared to the best track, the assimilation of inner-core aircraft observations by HEDAS produces generally robust analyses with respect to position and intensity. A mean position error of 38 km is obtained, and the position errors depend on intensity; in weaker cases, greater position errors are observed, likely both because the wind-position correlations are expected to be weaker because of weaker radial gradients and because the center finding process itself becomes more uncertain. For intensity, 10-m maximum wind speed exhibits a mean error of -1.1 m s^{-1} that is not statistically significant at the 95% confidence level because of the large uncertainty associated with this intensity parameter. Meanwhile, a much smaller variability in minimum sea level pressure renders the 3.7-hPa mean error (underestimation of intensity) statistically significant. The tangential wind speed also exhibits similar indications of statistically significant intensity underestimation through both its azimuthally averaged maximum (4.0 m s^{-1} mean error) and its wavenumber-0 amplitude (2.4 m s^{-1} mean error). Regardless, all of these indicators of intensity of the primary circulation exhibit strong correlations with their observed counterparts; linear regression analysis reveals coefficient of determination values in the range of 87%–97% for all cases considered.

HEDAS also produces good analyses in other aspects of the structure of the primary circulation, such as RMW and wavenumber-1 asymmetry. For RMW, the assimilation of inner-core observations performs the best in the hurricane cases. Obtaining a good analysis of wavenumber-1 asymmetry appears to be more challenging, but when 15% of the cases are excluded as outliers, a relatively good degree of agreement with observations is achieved. A tendency in HEDAS to axisymmetrize the primary circulation is also noted, albeit by about 5% of the total variance explained. Overall, a coherent picture emerges that suggests that inner-core, high-resolution observations, when assimilated by a state-of-the-art data assimilation technique such as the EnKF, result in analyses of the primary circulation that are realistic in terms of intensity, wavenumber-0 radial structure, as well as wavenumber-1 azimuthal structure, although the smaller signal-to-noise ratio in the wavenumber-1 component of the tangential wind leads to a weaker statistical signal in the quality of the analyses in this respect.

Estimation of the secondary circulation in HEDAS is found to be more challenging. Systematic underestimation of the maximum azimuthally averaged radial inflow and overestimation of the depth of the inflow layer are observed. It is believed that this is the result of a model bias in the experimental HWRF caused by the overdiffusive nature of the planetary boundary layer parameterization utilized. Further investigation of the impact of forecast biases on the final HEDAS analyses is presented in Vukicevic et al. (2013).

In terms of the thermodynamic structure, HEDAS overestimates both temperature and specific humidity for weaker cases and underestimates the warm-core perturbation in stronger cases. It should be noted here that the update of the thermodynamic variables in the current configuration of HEDAS relies more heavily on the indirect information content of the kinematic observations, as temperature observations that are assimilated are relatively sparse. Consequently, the observed errors in the analyzed thermal structure are expected to be caused by both an imbalance between the volume of the kinematic and thermodynamic observations in general and the suboptimal ensemble covariances between kinematic and thermodynamic fields.

Our future plans include addressing the former issue through observing system simulation experiments to explore the optimal combination of kinematic and thermodynamic information content in vortex-scale data assimilation. The latter issue is much more complex as it is linked to model error. Recently, upgrades to the operational HWRF model addressed model bias through adjustments in the vertical diffusion parameter in the boundary layer as well as momentum and heat exchange coefficients in the surface layer (S. G. Gopalakrishnan 2012, personal communication). Starting in the 2012 hurricane season, these model upgrades were integrated into HEDAS, which is expected to lead to a more realistic representation of the secondary circulation in

analyses. Work is also under way to investigate methods of accounting for model error in the HEDAS ensemble through the uncertainties associated with the subgrid-scale processes such as surface fluxes and boundary layer turbulence. Finally, methods of balancing solutions between the primary and secondary circulations to constrain the spinup of the vortex structure in short-range forecasts are being explored.

Acknowledgments. The authors acknowledge funding from the NOAA Hurricane Forecast Improvement Project (HFIP) that supported this work and provided the computing resources. This research was carried out (in part) under the auspices of CIMAS, a joint institute of the University of Miami and NOAA, Cooperative Agreement NA67RJ0149. HRD Director Dr. Frank Mark's guidance and leadership have been instrumental in the success of the HEDAS project. The commitment and effort of NOAA Aircraft Operations Center and Air Force Reserve flight crews in providing observations is greatly appreciated. Shirley Murillo and Bachir Annane from the HRD H*Wind group have provided the H*Wind data. Dr. Jeffrey Whitaker of NOAA/ESRL has provided the GFS/EnKF data. HRD internal reviewers Dr. Paul Reasor and Jason Dunion are also acknowledged for their constructive criticism to improve the manuscript. Dr. Jason Sippel and another anonymous reviewer have provided very valuable peer reviews to bring the manuscript to its final form.

REFERENCES

Aberson, S. D., 2009: 10 years of hurricane synoptic surveillance (1997–2006). *Mon. Wea. Rev.,* **138,** 1536–1549.

——, M. L. Black, R. A. Black, R. W. Burpee, J. J. Cione, C. W. Landsea, and F. D. Marks Jr., 2006: Thirty years of tropical cyclone research with the NOAA P-3 aircraft. *Bull. Amer. Meteor. Soc.,* **87,** 1039–1055.

Aksoy, A., D. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.,* **137,** 1805–1824.

——, ——, and ——, 2010: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble forecasts. *Mon. Wea. Rev.,* **138,** 1273–1292.

——, S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, 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. G. Gopalakrishnan, S. A. Michelson, F. D. Marks, and M. T. Montgomery, 2012: Impact of physics representations in the HWRFX model on simulated hurricane structure and wind–pressure relationships. *Mon. Wea. Rev.*, **140,** 3278–3299.

Berg, R. J., and L. A. Avila, 2011: Atlantic hurricane season of 2009. *Mon. Wea. Rev.,* **139,** 1049–1069.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.,* **132,** 1982–2005.

——, L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Wea. Rev.,* **139,** 272–294.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.,* **46,** 3077–3107.

Eastin, M. D., P. G. Black, and W. M. Gray, 2002: Flight-level thermodynamic instrument wetting errors in hurricanes. Part I: Observations. *Mon. Wea. Rev.,* **130,** 825–841.

Ek, M. B., and Coauthors, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.,* **108,** 8851, doi:10.1029/2002JD003296.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

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

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., 422–442.

Gao, J., M. Xue, A. Shapiro, and K. K. Droegemeier, 1999: A variational method for the analysis of three-dimensional wind fields from two Doppler radars. *Mon. Wea. Rev.,* **127,** 2128–2142.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723–757.

Gopalakrishnan, S. G., and Coauthors, 2002: An operational multiscale hurricane forecasting system. *Mon. Wea. Rev.,* **130,** 1830–1847.

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

——, Q. Liu, T. Marchok, D. Sheinin, N. Surgi, R. Tuleya, R. Yablonsky, and X. Zhang, 2010: Hurricane Weather and Research and Forecasting (HWRF) model scientific documentation. NOAA/NCAR/Development Test Center, 75 pp. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRF_final_2-2_cm.pdf.]

——, S. Goldenberg, T. Quirino, X. Zhang, F. Marks Jr., K.-S. Yeh, R. Atlas, and V. Tallapragada, 2012: Toward improving high-resolution numerical hurricane forecasting: Influence of model horizontal grid resolution, initialization, and physics. *Wea. Forecasting,* **27,** 647–666.

——, F. Marks, 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.

Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. J. Benjamin, 2011a: Predictions of 2010's tropical cyclones using the GFS

and ensemble-based data assimilation methods. *Mon. Wea. Rev.,* **139,** 3243–3247.

——, ——, D. T. Kleist, M. Fiorino, and S. J. Benjamin, 2011b: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.,* **139,** 668–688.

Hock, T. F., and J. L. Franklin, 1999: The NCAR GPS dropsonde. *Bull. Amer. Meteor. Soc.,* **80,** 407–420.

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

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

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796–811.

Landsea, C. W., and Coauthors, 2004: The Atlantic hurricane database re-analysis project: Documentation for the 1851–1910 alterations and additions to the HURDAT database. *Hurricanes and Typhoons: Past, Present and Future,* R. J. Murname and K.-B. Liu, Eds., Columbia University Press, 177–221.

Lorsolo, S., and A. Aksoy, 2012: Wavenumber analysis of azimuthally distributed data: Assessing maximum allowable gap size. *Mon. Wea. Rev.,* **140,** 1945–1956.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102** (D14), 16 663–16 682.

Moon, I.-J., I. Ginnis, T. Hara, and B. Thomas, 2007: A physics-based parameterization of air–sea momentum flux at high wind speeds and its impact on hurricane intensity predictions. *Mon. Wea. Rev.,* **135,** 2869–2878.

Moyer, A. C., J. L. Evans, and M. Powell, 2007: Comparison of observed gale radius statistics. *J. Meteor. Atmos. Phys.,* **97,** 41–55.

Nguyen, L. T., and J. Molinari, 2012: Rapid intensification of a sheared, fast-moving hurricane over the Gulf Stream. *Mon. Wea. Rev.,* **140,** 3361–3378.

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

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

Reasor, P. D., M. D. Eastin, and J. F. Gamache, 2009: Rapidly intensifying Hurricane Guillermo (1997). Part I: Low-wavenumber structure and evolution. *Mon. Wea. Rev.,* **137,** 603–631.

Rogers, R., S. Lorsolo, P. Reasor, J. Gamache, and F. Marks, 2012: Multiscale analysis of tropical cyclone kinematic structure from airborne Doppler radar composites. *Mon. Wea. Rev.,* **140,** 77–99.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.,* **131,** 1663–1677.

Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. *Wea. Forecasting,* **27,** 715–729.

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.

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.,* in press.

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.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913–1924.

——, and ——, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.,* **140,** 3078–3089.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.,* **132,** 1238–1253.

——, Y. Weng, J. A. Sippel, Z. Meng, and C. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.,* **137,** 2105–2125.

——, ——, 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, doi:10.1029/2011GL048469.

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

Zhang, X., K.-S. Yeh, T. S. Quirino, S. G. Gopalakrishnan, F. D. Marks Jr., S. B. Goldenberg, and S. Aberson, 2011: HWRFx: Improving hurriance forecasts with high-resolution modeling. *Comput. Sci. Eng.,* **13,** 13–21, doi:10.1109/MCSE.2010.121.