Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD’s HEDAS:
Evaluation of the 2008-2011 vortex-scale analyses

Altuğ Aksoy1,2, Sim D. Aberson2, Tomislava Vukicevic2,
Kathryn J. Sellwood1,2, and Sylvie Lorsolo1,2

1 Cooperative Institute for Marine and Atmospheric Studies, Univ. of Miami

2 Hurricane Research Division, NOAA/AOML

Miami, Florida

To be submitted to Monthly Weather Review

Last edited: 21 May 2012

*Corresponding author address*: Dr. Altuğ Aksoy, CIMAS, Rosenstiel School,
4600 Rickenbacker Cswy, Miami, FL 33149.

Email: aaksoy@rsmas.miami.edu

**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 domains. HEDAS is applied to 83 cases from years 2008-2011. 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 due to the over-diffusive nature of the planetary boundary layer parameterization utilized. Thermodynamic deviations from the observed structure are believed to be due to both an imbalance between the number of the kinematic and thermodynamic observations in general and the sub-optimal ensemble covariances between kinematic and thermodynamic fields. Future plans are discussed to address these challenges.

# 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 2009). Several factors could be contributing to this, including model deficiencies, suboptimal initialization due to the lack of observations in the peripheral environment as well as 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 aspect of the TC prediction problem and specifically explore the impact of TC inner-core observations on the analyses of high-resolution vortex structure using the ensemble Kalman filter (EnKF) technique.

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 1998). Recent success with assimilating radar observations of continental convective storms (e.g., 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 EnKF. In a proof-of-concept study, Zhang et al. (2009) demonstrated that the EnKF exhibited more skill than 3DVAR in predicting the rapid formation and intensification 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 2011; Zhang et al. 2011) and demonstrated improvement in the representation of the vortex structure in Hurricane Katrina (2005), as well as a reduction in intensity forecast error in 61 cases from 2008-2010 when compared to operational dynamical models.

The Hurricane Weather Research and Forecasting (HWRF; e.g., Rappaport et al. 2009) Hurricane Ensemble Data Assimilation System (HEDAS, Aksoy et al. 2012, to be referred to as A12 hereafter) is an ensemble-based data assimilation system developed at the National Oceanic and Atmospheric Administration (NOAA) Hurricane Research Division (HRD) to utilize airborne, high-resolution TC observations collected by NOAA’s WP-3D (P-3) aircraft (e.g., Aberson et al. 2006), high-altitude Gulfstream-IV (G-IV) jet (e.g., Aberson 2009), as well as the C-130 aircraft of the 53rd Weather Reconnaissance squadron of the U.S. Air Force Reserve Command (e.g., Rappaport et al. 2009). HEDAS comprises of an EnKF and HRD’s experimental HWRF model (Gopalakrishnan et al. 2011). A12 demonstrated the value of assimilating simulated airborne Doppler radar radial wind data in HEDAS and showed that radar wind observations not only had direct positive impact on the vortex wind structure in a TC but also indirect positive impact on the vortex 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 2008-2011 Atlantic hurricane seasons. The high-resolution vortex-scale analyses that are generated are evaluated for position, intensity, and structure in comparison to various observation platforms and analysis systems. 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.

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.

# The Real-Time HEDAS

## 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 2 two-way-interacting computational domains (see Table 1 for details). The vortex-following nest motion of the inner domain (Gopalakrishnan et al. 2002 and 2006) is suppressed during spin-up and data assimilation cycles and all ensemble members are initialized with co-located inner domains to facilitate grid-point-based spatial covariance computations in the EnKF. The 10°x10° inner nest is bigger than that of A12 to encompass the entire circulations 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. (2011).

## HEDAS

Similar to the data assimilation system described in Zhang et al. (2009) in its mechanics, 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 of Whitaker and Hamill (2002) are simplified following Snyder and Zhang (2003) equations 4-7. Three-dimensional, distance-dependent covariance localization, using a compactly supported fifth-order correlation function following Gaspari and Cohn (1999), is applied. Localization length scale 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 and 2011b). An ensemble spin-up is initialized 6 h prior to the synoptic time around which a respective NOAA P-3 flight is centered. The spin-up is carried out for 3-4 h (until the first observations are available) to develop appropriate covariance structures relevant for the scales at which the data assimilation is performed. The experimental GFS EnKF initial perturbations, without any covariance inflation, are found to result in comparable ensemble spread to that in A12. Therefore, unlike A12, no covariance inflation is applied in the real-data experiments discussed here.

# 2008-2011 aircraft cases considered

A total of 83 cases (20 TCs) are considered in which the NOAA P-3 tail Doppler radar collected data (see Table 2 for a list of the individual cases). All but two cases (2011 Hurricane Hilary in the eastern North Pacific) were sampled in the Atlantic basin. The geographical distribution of the observed positions for these cases, as obtained from the National Hurricane Center’s (NHC) HURDAT database (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 due to the range limitations of the NOAA P-3 aircraft. The distribution of the cases according to the 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 have intensities of tropical storm or category-1 hurricane (17.5-42 m s-1).

# Observations assimilated

The types of observations assimilated in HEDAS include Doppler radial wind super-observations (superobs, see A12 for details), GPS dropwindsondes (Hock and Franklin 1999, to be called dropsonde for brevity hereafter), aircraft flight-level wind and temperature, and Stepped Frequency Microwave Radiometer (SFMR, Uhlhorn et al. 2007) 10-m wind speed measurements. Further details on assimilated observations and their processing can be found in Table 3. Beginning with the first observation available within the model inner nest, observations are grouped and assimilated in 1-h assimilation windows according to their sampling time. The number of cases as a function of number of assimilation cycles processed in this manner 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 Air Force Reserve C-130 flights and/or long “on station” times for storms that are closer to aircraft bases/deployment sites.

The distribution of the number of cases as a function of number of observations assimilated for each observation platform processed (Fig. 3, top histograms in each panel) reveals that the Doppler wind observations significantly outnumber observations from other platforms (by one order of magnitude in the case of flight-level observations and by two orders of magnitude in the case of dropsonde and SFMR observations). When broken down by intensity category also (Fig. 3, 2-d matrix plots in each panel), stronger cases generally have more observations assimilated for each platform. The modes of these distributions also shift toward higher values in each case, except for dropsonde observations that have the greatest number of cases in the smallest observation bin for each intensity category. These differences can be explained by a combination of the following facts: (a) More Doppler wind observations are sampled in stronger storms as stronger convection results in greater areal and vertical hydrometeor coverage for return signal. (b) More flight-level and SFMR observations are obtained in stronger storms as the likelihood of occurrence of overlapping NOAA P-3 and Air Force Reserve C-130 flights increases with storm intensity. (c) The number of dropsondes that can be launched from both the NOAA and Air Force Reserve aircraft is limited due to logistical considerations and does not depend on storm intensity.

When the number of observations assimilated per case is plotted by height
(Fig. 4), broader distributions of flight-level observations (Fig. 4a) with secondary maxima at lower altitudes are apparent for weaker cases. This results from the combination of a greater variability of flight altitude in weaker cases as well as some very-low-altitude Air Force Reserve C-130 flights in the weakest cases. In stronger cases, the distributions become unimodal and peak near 3 km altitude, which is the standard Air Force Reserve C-130 and NOAA P-3 flight altitude (near 10,000 ft) 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 than those of flight-level observations, indicating that such peaks occur below the aircraft. Finally, the peaks of the dropsonde observation populations also shift upward with intensity from near the surface in weaker storms, which results from the generally lower NOAA P-3 and Air Force Reserve C-130 aircraft altitudes in weaker storms. Observations sampled above ~3-4 km altitude are from NOAA G-IV dropsondes.

# Results

## Observation space diagnostics

In this section, the focus is on innovation-based statistics, where innovations represent observation-minus-model (forecast or analysis) differences.

### Frequency distributions of observation innovations

Figure 5 shows the frequency distributions of various types of observations that were investigated to understand the overall statistical behavior of the analyses in direct comparison to observations. For each observation type assimilated by HEDAS, probability distributions of prior observation innovations (i.e., observation-minus-forecast differences) at the first assimilation cycle are shown with light gray bars, while those of posterior observation innovations (i.e., observation-minus-analysis differences) at the last assimilation cycle are shown with dark gray bars. First-cycle prior innovations reflect the observation-background differences prior to the onset of data assimilation, while last-cycle posterior innovations reflect the observation-analysis differences after the completion of assimilation of all inner-core data. 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 relative to the storm center, 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 ~0 m s-1. Meanwhile, the wider probability distribution of the first-cycle prior innovations clearly points to greater wind errors carried over from the lower-resolution initial GFS-EnKF ensemble. The analysis errors appear to be significantly reduced by the time the final assimilation cycle is reached, as deduced from the much narrower innovation distribution of Doppler wind observations: There is a 46% likelihood to encounter an absolute Doppler wind 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 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.

Unlike Doppler wind, SFMR observations carry information about the absolute wind speed at 10 m. Therefore, SFMR innovations are more likely to reflect model/analysis biases of storm intensity. Indeed, the probability distribution of the first-cycle prior SFMR innovations peaks between 1-3 m s-1, indicating that the background 10-m wind speed is systematically under-predicted. An alternative way of looking at this is to measure the total probability of under-prediction vs. over-prediction by the first-cycle background: The likelihood of encountering under-predicted 10-m wind speed is 68% as opposed to 32% for over-prediction. Data assimilation helps to reduce this apparent bias and symmetrizes the posterior distribution of innovations in the final analysis: The likelihood of encountering under-predicted 10-m wind speed is now 57%, suggesting that observation-analysis differences have become more random in nature, but still suffer from a slight bias of under-estimation of intensity.

There also appears to be a slight easterly (positive) bias (57% likelihood) for the first-cycle prior innovations of zonal wind speed (combined flight-level and dropsonde platforms). However, since zonal wind speed observation is a vector, this result cannot be directly linked to under-prediction of intensity. (In other words, strong intensity is equally likely to result in positive, eastward, and negative, westward, zonal wind speed.) Rather, a bias in zonal wind speed is more likely to result from systematic storm track errors. Meanwhile, the bias is mostly removed in the final analysis: The likelihood of encountering under-estimated easterly wind is down to 52%. This suggests that aircraft tropical cyclone observations have the potential to impact the inner core as well as the environment of a tropical cyclone. A smaller but negative (southerly) bias also exists in the meridional wind speed in the first-cycle prior innovations (not shown). The results for the meridional wind speed 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 under-predicted. This suggests a weaker-than-observed background vortex available to the EnKF in the first assimilation cycle. In the final analysis, the magnitude of this apparent bias is reduced. However, the probability distribution of the posterior innovations now reveals a 53% likelihood of negative bias.

### Vertical variation of observation innovations

Figures 6-8 show the vertical variation of observation innovation statistics for Doppler wind speed, SFMR, temperature, and zonal wind speed observations as computed before the first assimilation cycle (i.e., first-cycle prior) and after the last assimilation cycle (i.e., last-cycle posterior) to illustrate how the assimilation of high-resolution aircraft observations impacts inner core structure. 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. 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 higher levels of the troposphere (~4-14 km), which is reduced in the analyses at the last cycle. Meanwhile, 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 an indication that the background wind field, even after a spin-up of 3-4 h from the global GFS/EnKF analysis ensemble, contains storms that are weak in intensity compared to observations. 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 assimilating inner-core observations and obtaining a good fit to observations in observation space.

Innovation statistics of SFMR observations depict a somewhat different picture than Doppler wind speed (Fig. 6, square symbols plotted at 0-m height in each panel) as mean and RMS background-observation departures are comparable in magnitude. Similar to Doppler wind speed, departures still increase with increasing storm intensity and approach ~10 m s-1 for major hurricane cases. Meanwhile, in the final analysis, mean innovations are reduced to below 1 m s-1 for all intensity categories and RMS innovations are reduced to ~4 m s-1, which is roughly the mean observation error for the SFMR platform[[1]](#footnote-1).

Although somewhat noisy in nature, innovation statistics for the zonal wind speed observations (as measured at flight level and by dropsondes) are similar to the Doppler wind speed and SFMR observations (Fig. 7), in the sense that the magnitude of the total (mean plus RMS) departures of the background from observations again exhibit positive correlation to storm intensity and approach ~10 m s-1 for major hurricane cases. 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, 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 at flight level and 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 mid-level 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 ~0.5 K, comparable in magnitude to observation error. However, stronger 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 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 here. Therefore, it is only logical to infer that what is lost in 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 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.

## Position and intensity compared to best track

In this study, verification of the HEDAS final analyses (to be referred to as “analyses” hereafter) against the best track is carried out at the nearest synoptic time to the final analyses. As the time of the final analyses is somewhat arbitrary and depends on when the respective NOAA flight ends, up to a 3-h difference is possible between verification and final analysis times. This introduces a slight over-estimation of analysis errors against the best track, especially for storm position, which tends to be temporally progressive in nature.

The distribution of the storm centers in analyses is compared to the best track storm centers at respective verification times (Fig. 9a). For a relevant storm-relative comparison, the analysis-best track displacements are shown with respect to observed direction of storm motion. 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 radial 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 to 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: (a) 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 correlation signals between wind observations and position. (b) The procedure of center finding itself is easier (and therefore more accurate) in stronger storms, mainly due to 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 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., under-estimation of intensity), although it is not statistically significant at the 95% confidence level. An even better linear regression fit is achieved for minimum sea-level pressure (97% variance explained, Fig. 10b). However, this is now accompanied by a more distinct, positive bias of 3.7 hPa (under-estimation of intensity) that is statistically significant. The analysis-best track similarities in maximum intensity and minimum sea-level pressure 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 82-83% variance explained. The separation between the two regression lines reflects the positive bias in analysis minimum sea-level pressure, although it is not statistically significant at the 95% confidence level (not shown).

## Primary circulation compared to airborne Doppler radar observations

The focus is now turned to the properties of the azimuthally averaged vortex structure in the analyses. Here, the comparison is carried out against observation-based vortex properties obtained from airborne Doppler radar composite analyses[[2]](#footnote-2)2. In terms of the primary circulation, maximum tangential wind speed (at any altitude) is investigated first. Figure 11a shows a robust linear fit between analyses and radar observations with 89% of the variance explained. 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 under-estimate 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 tropical cyclone intensity.

The dependence of the analysis intensity difference from observed values 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 tangential wind speed. It is evident that the mode of the probability distribution shifts toward more negative values (more under-estimation) with increasing observed intensity. Furthermore, the widening distributions with increasing intensity point to the greater degree of disagreement between HEDAS and radar observations on tropical cyclone intensity. Possible causes of such systematic errors in the HEDAS analyses are investigated in Vukicevic et al. (2012).

Analyses are also compared against how well they represent the structure of the primary circulation. In Fig. 12a, this comparison is carried out for the radius of maximum azimuthally averaged tangential wind speed (RMW) at 1-km altitude, where the tangential wind speed is typically maximized. When all cases are considered, although a positive linear relationship between HEDAS and radar observations is evident, the analyses explain only 39% of the variance of the RMW in radar observations (dashed line). However, a much better agreement is achieved (62% variance explained) when only hurricane 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 speed is further investigated through the azimuthal phase of the wavenumber-1 asymmetry at 1-km altitude (Fig. 12b). When all cases are considered, almost no linear statistical relationship is discernible between analyses and radar observations (3% variance explained, dashed line). Unlike RMW, wavenumber-1 asymmetry correlation between HEDAS and observations does not appear to respond to intensity either: for only the hurricane cases, the variance explained remains very low at ~1% (not shown). However, an interesting observation is made that 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, 15 cases (24% of all cases) are deemed outliers. When regression analysis is performed excluding these outlier cases, a much better agreement between analyses and observations is achieved (79% variance explained, 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 further examined in terms of the linear regression between analyses and radar observations of azimuthal wavenumber 0 and 1 components (Fig. 13). All of the statistics here are computed at 1 km altitude. The magnitude of the wavenumber-0 amplitude (Fig. 13a) reveals a very good linear fit at 93% variance explained. A statistically significant negative bias of 2.4 m s-1 is also apparent, indicating that analyses tend to slightly under-estimate the intensity of the mean azimuthally averaged tangential wind speed. These results are consistent 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 is also compared between analyses and radar observations (Fig. 13b). A situation that is similar to that discussed for the phase of wavenumber-1 tangential wind speed azimuthal asymmetry (Fig. 12b) is obtained: When all cases are considered, the linear fit between analyses and radar observations is poor (R2 = 11%). On the other hand, when outlier cases are not considered (an outlier here occurs when the absolute difference between an analysis and observation is 30% or greater), the linear fit is improved and
R2 = 70% is obtained. At the 30% threshold level, only 12 outlier cases are identified (19% of all cases considered). A similar scenario is also encountered for the variance explained by the wavenumber-1 component of the tangential wind speed (Fig. 13d), where the 30% threshold level results in 12 outlier cases, without which a good linear fit (R2 = 67%) is obtained between analyses and radar observations. 11 of the 12 outlier cases are found to be the same between wavenumber-0 and wavenumber-1 variance explained parameters, indicating that there exists a common underlying cause for outlier behavior between them. Further reinforcing this line of thinking is the fact that wavenumber-0 and wavenumber-1 components of the tangential wind speed together consistently account for more than 90% of the total variance between analyses and radar observations: On average, the total variance explained by wavenumber-0 and wavenumber-1 components is 94% and 96% for analyses and radar observations, respectively.

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 R2=40%, 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. However, considering that the average wavenumber-1 amplitude magnitude, when outlier cases are not considered, of 3.3 m s-1 for analyses (2.9 m s-1 for radar observations) is comparable to the standard error of the wavenumber-0 amplitude of 2.1 m s-1 (2.3 m s-1 for radar observations) itself, the wavenumber-1 amplitude clearly exhibits a low signal-to-noise ratio compared to the variability of its wavenumber-0 counterpart.

## Secondary circulation compared to radar observations

Estimating the secondary circulation with HEDAS is found to be more challenging. Figure 14 summarizes this finding by comparing 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 under-estimate the magnitude of the radial inflow by 8.9 m s-1 and over-estimate the depth of the radial inflow by 0.8 km. Both estimates are statistically significant at the 95% confidence level. Bao et al. (2012) documented the systematic positive inflow depth bias in the experimental HWRF model (compare their Fig. 4a, which represents the experimental HWRF configuration, to their Fig. 4b) that arises from the choice of the physical parameterizations. It is believed that the findings here are consistent with their findings: The depth of the inflow layer is systematically over-estimated in HEDAS as a result of an existing model bias. 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.

## 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 H\*Wind analyses at the surface (Powell et al. 1998).

Composite radial profiles are computed as follows: For both observations (gray lines) and analyses (black lines), first, azimuthal averages are computed on a case-by-case basis (normalized in radial distance by respective RMW) and composites are then obtained for the respective intensity criteria. Flight-level observations are obtained from in-situ sensors on the P-3 aircraft. Individual observations report geopotential height values that are then converted to pressure and averaged for all cases to yield the average composite altitude shown by “FL Ob” in figure inserts. For analyses, continuous fields are converted to polar coordinates and sampled at 60º azimuthal resolution (to mimic azimuthal sampling during actual P-3 flights) at reported mean pressure altitude of the flights. Composite altitudes, shown in figure inserts as “Anlys”, are computed as averages of such mean pressure altitudes for the cases involved in each panel. Mean RMW (at flight level for Figs 15-17 and at surface for Fig. 18), as computed from observations and analyses, is also shown in the figure inserts for reference.

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

At flight level, the structure of horizontal wind speed is investigated first (Fig. 15). The mean structures are very similar for all cases up to hurricane category 2 (Fig. 15a-b) and statistically not distinguishable at 95% confidence level. For stronger hurricanes (Fig. 15c), analyses under-estimate peak wind speed by ~5 m s-1. It is also interesting to note that, except for weaker cases outside RMW, the statistical uncertainty of wind speed at flight level is comparable between observations and analyses.

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

The systematic over-estimation of temperature and specific humidity in analyses is likely due to a combination of three factors. First, the update of the thermodynamic fields in HEDAS relies heavily on the indirect wind speed information obtained from the numerous Doppler wind speed observations and on the quality of correlations between wind speed and thermodynamic fields in the ensemble. Although flight-level and dropsonde temperature observations are assimilated, they are much more sparse and their impact appears to be limited to the cyclone core, likely due to the better azimuthal resolution achieved in the core. Outside the core, limited impact from temperature observations reveals more clearly the underlying systematic tendencies in HEDAS (Vukicevic et al. 2012). In addition to the possible ensemble sampling issues, the systematically deeper boundary layer that was discussed in section 5d is also expected to render a warm and moist bias in the boundary layer. Meanwhile, it is interesting that these biases become much less pronounced outside the core for major hurricanes despite persistent boundary layer depth biases. The distinct under-estimation of the warm core for these cases, however, appears to be consistent with the under-estimation of kinematic intensity.

Finally, at the surface, analyses are in good general agreement with H\*Wind analyses (Fig. 18). Under-estimation of peak wind speed by ~2 m s-1 for major hurricanes in analyses is consistent with previous findings. Furthermore, in analyses, storm size is somewhat smaller than that in H\*Wind analyses, especially for hurricanes.

# Summary and discussion

The impact of inner-core tropical cyclone observations from aircraft on analyses of high-resolution vortex structure is investigated through the use of NOAA/AOML/HRD’s EnKF-based data assimilation system, HEDAS. 83 cases (20 individual tropical cyclones) from the years 2008-2011 are considered. With the exception of two Hilary (2011) cases in the eastern North Pacific basin, all cases are observed in the Atlantic basin. Observed intensity ranges from tropical depression to category-4 hurricane, with most cases falling into tropical storm category.

HEDAS assimilates available observations from NOAA P-3, G-IV, and/or Air Force Reserve flights in one-hour intervals. Among the 83 cases processed, 5 assimilation cycles (4 hours of cycling) is the most frequently encountered length of assimilation period. Assimilated observation types include aircraft Doppler radar wind superobs, temperature and zonal/meridional wind speed from aircraft flight level measurements and deployed dropsondes, and SFMR 10-m wind speed retrievals. Among these observation types, Doppler wind superobs are the most numerous and comprise about one order of magnitude greater number of observations assimilated than the other types of aircraft data (flight-level, dropsonde, and SFMR). Conversely, the number of Doppler wind superobs assimilated exhibits a greater dependence on intensity, with more observations assimilated in strongest hurricanes, while other observation types exhibit much smaller dependence on intensity.

Observation space diagnostics reveal some under-dispersiveness in the HEDAS ensemble, especially for Doppler wind speed observations. Although this is a similar finding to A12, 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 here. Therefore, it is inferred that what is lost in 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 due to 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 by A12, where the application of various covariance inflation techniques reduced but not totally eliminated 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, which exhibits dependence on intensity: In weaker cases, greater position errors are observed, likely both because the wind-position correlations are expected to be weaker due to weaker radial gradients and because center finding itself becomes more uncertain. As for intensity, 10-m maximum wind speed exhibits a mean error of -1.1 m s-1 that is not statistically significant at 95% confidence level due to 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 (under-estimation of intensity) statistically significant. The tangential wind speed also exhibits similar indications of statistically significant intensity under-estimation 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 variance explained values that vary in the range of 87-97% for all cases considered.

HEDAS produces good analyses also 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 with the exception of about 15-20% of the cases, a relatively good fit to observations is achieved (40-79% variance explained depending on the parameter). 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 under-estimation of the maximum azimuthally averaged radial inflow and over-estimation 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 due to the over-diffusive 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. (2012).

Recently, upgrades to the HWRF operational model addressed this 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). Beginning from the 2012 hurricane season, these model upgrades will be adopted in HEDAS, which is expected to lead to a more realistic representation of the secondary circulation in analyses.

In terms of the thermodynamic structure, HEDAS over-estimates both temperature and specific humidity for weaker cases and under-estimates 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 due to both an imbalance between the volume of the kinematic and thermodynamic observations in general and the sub-optimal 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. Upgrading HEDAS, for the 2012 hurricane season, to the updated HWRF configuration with improved surface layer and boundary layer parameterizations is expected to have positive impact on the quality of analyses. Work is also underway to investigate ways to account for model error in the HEDAS ensemble through the uncertainties associated with the sub-grid-scale processes such as surface fluxes and boundary layer turbulence. Finally, methods of balancing solutions between the primary and secondary circulations to constrain the spin-up of the vortex structure in short-range forecasts are being explored.

# Acknowledgements

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 and Air Force Reserve flight crews in providing observations is greatly appreciated. The HRD modeling group, especially Dr. Xuejin Zhang and Dr. Thiago Quirino, has provided valuable support for the experimental HWRF. 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.

# References

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

Aberson, S. D., 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., 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.*, in print.

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.

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.*, accepted.

Berg, R. J. and L. A. Avila, 2011: Annual summary: 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.

Dowell, D. C., 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 (D22)**, 8851.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10143-10162.

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

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

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

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.

Hamill, T. M., J. S. Whitaker, 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., amd H.-L. Pan, 1998: Convective trigger function for a mass-flux cumulus parameterization scheme. *Mon. Wea. Rev.*, **129**, 1164-1178.

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

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., C. Anderson, N. Charles, G. Clark, J. Dunion, J. Fernandez Partagas, P. Hungerford, C. Neumann, and M. Zimmer, 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.

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)**, 16663-16682.

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.

Powell, M. D., S. H. Houston, L. R. Amat, and N. Morisseau-Leroy, 1998: The HRD real-time hurricane wind analysis system. *J. Wind Engineer. Ind. Aerody.*, **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.

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

Uhlhorn, E. W., P. G. Black, J. L. Franklin, M. Goodberlet, J. Carswell, and Al. 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, 2012: Impact of short-term dynamical forecast bias in ensemble Kalman filter data assimilation of high-resolution tropical cyclone observations. To be submitted to *Mon. Wea. Rev.*

Weng, Y. and F. Zhang, 2011: Assimilating airborne Doppler radar observations with an ensemble Kalman filter for convection-permitting hurricane initialization and prediction: Katrina (2005). *Mon. Wea. Rev.*, accepted.

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

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.

Zhang, F., 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.

Zhang, F., Y. Weng, 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.*, doi:10.1029/2011GL048469, in press.

# Figure captions

**Figure 1.** (a) Geographical distribution of the case position as observed in the best track dataset. (b) Frequency distribution of the case intensity category as observed in the best track dataset. Cases that do not exist in the best track database (see Table 2) are assigned “<TS” category. “TS” and “H” stand for tropical storm and hurricane, respectively.

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

**Figure 3.** Distribution of the number of cases according to the number of observations assimilated per platform and intensity category (2-d matrix plots in each panel, color scale on top right), cumulative according to number of observations per platform (top histograms in each panel), and cumulative according to intensity category (bottom right histogram). The 2-d matrix plots are normalized by the respective maxima of populations for each intensity category. “TS”, “H”, and “HM” stand for tropical storm, hurricane, and major hurricane, respectively.

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

**Figure 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 zonal temperature observations. 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.

**Figure 6.** Observation-space innovation (observation-minus-model) statistics for Doppler wind speed and SFMR (shown with squares at 0-m height) the first-cycle prior (thick gray) and final-cycle posterior (thick black) distributions. Left panels: mean innovations, center panels: RMS innovations, and right panels: spread sufficiency ratio. For spread ratio, the black lines represent final-cycle prior distributions. The statistics are aggregated for weaker-than-tropical-storm (first row), tropical storms (second row), hurricanes of category 1 and 2 (third row), and major hurricanes (fourth row). Thin lines represent 95% confidence intervals.

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

**Figure 8.** As in Figure 6, but for temperature observations.

**Figure 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 (2-d matrix plot, color scale on right), and cumulative as a function of the analysis-observed radial distance of storm centers (histogram). The 2-d matrix plot population bin values are normalized by the respective maxima of populations for each intensity category.

**Figure 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 thick line represents the linear regression between analysis and observations. The coefficient of determination (R2) is presented in the lower-right box. The mean analysis-observation difference along with its 95% confidence interval bounds is given in the upper-left box. The dashed gridlines represent intensity category thresholds. (b) As in (a), but for minimum sea-level pressure. (c) Wind-pressure relationship in observed (square markers, solid linear regression line) and analysis (diamond markers, dashed linear regression line) data.

**Figure 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 thick line represents the linear regression between analysis and observations. The coefficient of determination (R2) is presented in the lower-right box. The mean analysis-observation difference along with its 95% confidence interval bounds is given in the upper-left box. (b) Number of cases as a function of analysis-observation difference and intensity category (2-d matrix plot, color scale on right), and cumulative as a function of analysis-observation difference (histogram). The 2-d matrix plot population bin values are normalized by the respective maxima of populations for each intensity category.

**Figure 12.** (a) Scatter diagram of analysis vs. radar-observed radius of maximum azimuthally averaged tangential wind speed (RMW, km) at 1-km altitude for all cases in the best track database. Filled squares represent hurricanes. The thick dashed (solid) line represents the linear regression between analyses and observations for all (hurricane) cases. The coefficients of determination (R2) are presented in the lower-right box. The mean analysis-observation differences along with their 95% confidence interval bounds are given in the upper-left box. (b) As in (a), but for the azimuthal phase of wavenumber-1 asymmetry (º from storm motion) at 1-km altitude. Filled squares represent cases that are deemed not to be outliers (see text). The thick dashed (solid) line represents the linear regression between analysis and observations for all (non-outlier) cases.

**Figure 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. (a) Wavenumber 0 amplitude (m s-1). (b) Wavenumber 0 variance explained (%). (c) Wavenumber 1 amplitude (m s-1). (d) Wavenumber 1 variance explained (%). The thick lines represent the linear regressions between analysis and observations for each parameter. In (b-d), filled squares represent cases that are deemed not to be outliers (see text). The dashed lines represent the linear regressions between analysis and observations for non-outlier cases.

**Figure 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 thick solid line represents the linear regression between analyses and observations for all cases. (b) As in (a), but for the depth of the radial inflow (km).

**Figure 15.** Composite radial profiles of azimuthally averaged horizontal wind speed at P-3 flight level for HEDAS final analyses (thick black) and in-situ aircraft observations (thick gray). 95% confidence intervals are shown as dashed lines. Average RMW and flight altitude (“Alt”) in analyses as well as observations are shown in the box inserts, along with corresponding 95% confidence intervals. (a) All cases that had tropical-storm or weaker intensity in the best track database. (b) All cases with category-1 and category-2 hurricane intensity. (c) Major hurricanes.

**Figure 16.** As in Fig. 15 but for temperature (K).

**Figure 17.** As in Fig. 15 but for specific humidity (g kg-1).

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

# List of tables

**Table 1.** Summary of the experimental setup.

**Table 2.** Summary of the cases considered.

**Table 3.** Observation platforms/types assimilated.

1. 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. [↑](#footnote-ref-1)
2. 2 Three-dimensional analyses of NOAA P-3 tail Doppler radar wind data (Gamache 1997, Gao et al. 1999, Reasor et al. 2009) are routinely performed aboard the aircraft after each eye penetration and transmitted to ground operations. For completeness, any analysis not performed and transmitted during the flight was made afterward using the same configuration to mimic what would have been done on the aircraft. After the completion of all analyses for a particular flight, they are composited in a storm-relative framework: for each analysis point, that with the highest wind speed is found, and the zonal and meridional components of the wind are used. [↑](#footnote-ref-2)