Assimilation of High-Resolution Tropical Cyclone Observations with an Ensemble Kalman Filter Using HEDAS: Evaluation of 2008–11 HWRF Forecasts

SIM D. ABERSON

NOAA/AOML/Hurricane Research Division, Miami, Florida

ALTUĞ AKSOY AND KATHRYN J. SELLWOOD

University of Miami/Cooperative Institute for Marine and Atmospheric Studies, Miami, Florida

TOMISLAVA VUKICEVIC

NOAA/AOML/Hurricane Research Division, Miami, Florida

XUEJIN ZHANG

University of Miami/Cooperative Institute for Marine and Atmospheric Studies, Miami, Florida

(Manuscript received 25 April 2014, in final form 20 October 2014)

ABSTRACT

NOAA has been gathering high-resolution, flight-level dropwindsonde and airborne Doppler radar data in tropical cyclones for almost three decades; the U.S. Air Force routinely obtained the same type and quality of data, excepting Doppler radar, for most of that time. The data have been used for operational diagnosis and for research, and, starting in 2013, have been assimilated into operational regional tropical cyclone models. This study is an effort to quantify the impact of assimilating these data into a version of the operational Hurricane Weather Research and Forecasting model using an ensemble Kalman filter. A total of 83 cases during 2008–11 were investigated. The aircraft whose data were used in the study all provide high-density flight-level wind and thermodynamic observations as well as surface wind speed data. Forecasts initialized with these data assimilated are compared to those using the model standard initialization. Since only NOAA aircraft provide airborne Doppler radar data, these data are also tested to see their impact above the standard aircraft data. The aircraft data alone are shown to provide some statistically significant improvement to track and intensity forecasts during the critical watch and warning period before projected landfall (through 60 h), with the Doppler radar data providing some further improvement. This study shows the potential for improved forecasts with regular tropical cyclone aircraft reconnaissance and the assimilation of data obtained from them, especially airborne Doppler radar data, into the numerical guidance.

1. Introduction

Numerical weather prediction is hampered by model deficiencies, suboptimal model initialization, and the inherent unpredictability of the system to be forecast. Though nothing can be done about the chaotic nature of the atmosphere, improvements to models and their initialization hold promise. Specific to tropical cyclone (TC) forecasting, efforts have been made to improve model forecast systems, including their data assimilation, under the aegis of the Hurricane Forecast Improvement Project (HFIP; Gall et al. 2013). As part of this project, assimilation of all high-resolution inner-core observations from airborne platforms (Aberson et al. 2006) is attempted for the first time.

The present study focuses on the impact of airborne observations on TC forecasts using the ensemble Kalman filter (EnKF). The EnKF is an advanced data assimilation technique utilizing short-range ensemble forecasts

Corresponding author address: Sim D. Aberson, NOAA/ AOML/Hurricane Research Division, 4301 Rickenbacker Causeway, Miami, FL 33149.
E-mail: sim.aberson@noaa.gov

to estimate flow-dependent spatial and cross correlations (Evensen 1994; Houtekamer and Mitchell 1998). F. Zhang et al. (2011) tested the assimilation of airborne Doppler radar wind data using an EnKF and the Advanced Research version of the Weather Research and Forecasting (ARW) Model. They demonstrated improvement in intensity forecasts in 61 cases during 2008–10 when compared to operational dynamical models, showing that reducing the uncertainty in mesoscale initial conditions could have a positive impact on TC forecast skill. This encouraging result suggests that improving the vortex-scale initial conditions with advanced data assimilation technology has the potential for improving TC intensity forecasts.

The Hurricane Weather Research and Forecasting (HWRF) version of the Hurricane Ensemble Data Assimilation System (HEDAS; Aksoy et al. 2012) is an EnKF-based data assimilation system; the current study utilizes high-resolution TC observations collected by NOAA's WP-3D (P3) aircraft (Aberson et al. 2006) and high-altitude Gulfstream-IV (G-IV) jet (Aberson 2010), as well as U.S. Air Force Reserve C-130 aircraft (Aberson et al. 2010), although HEDAS has the capability of assimilating any observation for which operators exist. Aksoy et al. (2013) demonstrated the value of assimilating airborne Doppler radar radial wind data with HEDAS on 83 cases in 20 TCs during the 2008–11 seasons. They found that the assimilation of high-resolution airborne observations results in realistic analyses of the primary circulation in terms of intensity, wavenumber-0 radial structure, and wavenumber-1 azimuthal structure. They showed direct positive impact on the vortex wind structure, but also indirect positive impact on the thermodynamic structure. Their comparison of the HEDAS analyses with independent Doppler-based wind analyses revealed a significant low intensity bias as well as a large underestimate of the low-level radial inflow for all cases. Vukicevic et al. (2013) found that these errors were caused by a short-term spindown of the entire vortex circulation in each data assimilation cycle, and that the magnitude was correlated with the TC intensity. These systematic errors were caused by both the radial and vertical components of the secondary circulation being much weaker than expected in the short-term forecasts that make up the HEDAS background. This was shown to be due to three main factors: the observations had limited information about the secondary circulation, the model has a bias toward rapid development of a too-deep planetary boundary layer, and the vertical velocity is not updated in the assimilation cycling because of the HWRF nonhydrostatic governing equations. These errors are likely to limit the impact of the data assimilation on subsequent forecasts, especially those for intensity.

This study extends the earlier ones by looking at the impact of the data assimilation on HWRF track, intensity, and basic structure forecasts using a variety of metrics. HEDAS and the version of HWRF used in the study are briefly described in section 2. The various forecast verifications are presented in section 3, with conclusions following.

2. Model and model initialization

Initial conditions are taken from the HEDAS analyses described in Aksoy et al. (2013) wherein the details of the methodology are presented. Briefly, HEDAS is based on a serial implementation of the square root EnKF (Whitaker and Hamill 2002). Further technical details are explained in Aksoy et al. (2012, 2013). In this study, 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) (Hamill et al. 2011a,b). An HWRF ensemble spinup is initialized 6 h prior to the synoptic time around which a NOAA P3 flight is centered. The spinup is carried out until the first observations are available (usually 3–4 h) to develop the covariance structures relevant for the scales at which data assimilation is performed. The data assimilation itself is carried out until the time of the last airborne Doppler radar observation, generally 3–5 h after the end of the spinup period. Further details can be found in Aksoy et al. (2013).

The research version of HWRF applied here (Gopalakrishnan et al. 2012; Yeh et al. 2012; X. Zhang et al. 2011) is configured with two two-way-interacting computational domains consisting of an outer domain and a vortex-following $10° \times 10°$ inner domain with 9- and 3-km horizontal grid spacings. All cases are in the Atlantic basin, and the outer mesh is configured as in the then-current operational version of HWRF except for the difference in resolution. Since assimilation is carried out on the inner domain only, all ensemble members are initialized with the domain collocated and with motion suppressed during the data assimilation to facilitate gridpoint-based spatial covariance computations. A detailed comparison of the physics parameterizations used in the experimental and operational HWRF can be found in Gopalakrishnan et al. (2012) and Yeh et al. (2012).

Analyses are made only when airborne Doppler radar observations were available. A total of 83 such cases from 20 individual TCs (Aksoy et al. 2013, their Table 2) are considered. The ensemble mean at the last cycle time (after which Doppler radar data are no longer available) is used to initialize the deterministic forecast; thus, if the final HEDAS cycle is at 0200 UTC, the model

is initialized with the analysis at that time. The operational GFS is used for initial and boundary conditions for the outer domain. The GFS fields are removed in the 3-km inner domain, and the HEDAS analysis is inserted. No special meshing at the domain boundaries is done; examination of the fields suggests that any discontinuities are small, and gravity waves are of the same magnitude just after initialization as those caused by mesh moves.

NOAA has been gathering high-resolution flight-level (pressure, temperature, humidity, and wind velocity, as well as surface wind speed from the stepped-frequency microwave radiometer), dropwindsonde (pressure, temperature, humidity, and wind velocity), and airborne Doppler radar data in TCs for almost three decades (Aberson et al. 2006). The characteristics of the data processing, including observation error, for these cases are presented in Aksoy et al. (2013, their Table 3). Until recently, only dropwindsonde data have been operationally assimilated into numerical models. This study represents the first test of the assimilation of high-resolution aircraft data into a version of HWRF. Three sets of model runs are considered. The first set is initialized with HEDAS analyses incorporating all data including the Doppler radar data. The second set is identical except that the airborne Doppler radar data are excluded from the assimilation system. A control set using the Geophysical Fluid Dynamics Laboratory model initial vortex (Kurihara et al. 1993, 1995) procedure is used for comparison (X. Zhang et al. 2011; Gopalakrishnan et al. 2012; Yeh et al. 2012).

3. Results

a. Track

The homogeneous track verification is shown in Fig. 1. Only those cases in which a TC exists in the postprocessed best track and is also identified by the model vortex tracker are compared. The great-circle distance between the two locations (the error) is calculated for each pair, and average errors are computed. The number of cases decreases with forecast time so that fewer than half the runs are verifiable by 4 days into the forecasts. Fewer than 83 runs are verified at 12 h due to either the TC not being in the best track or not being trackable by the model at that time. No 0-h verification is available, since the model runs were initialized 1–3 h after the synoptic time; a detailed investigation of the analysis quality can be found in Aksoy et al. (2013). Forecast errors decrease from 108 to 120 h into the forecasts due to the small sample sizes at those forecast times. No large difference between the forecast track
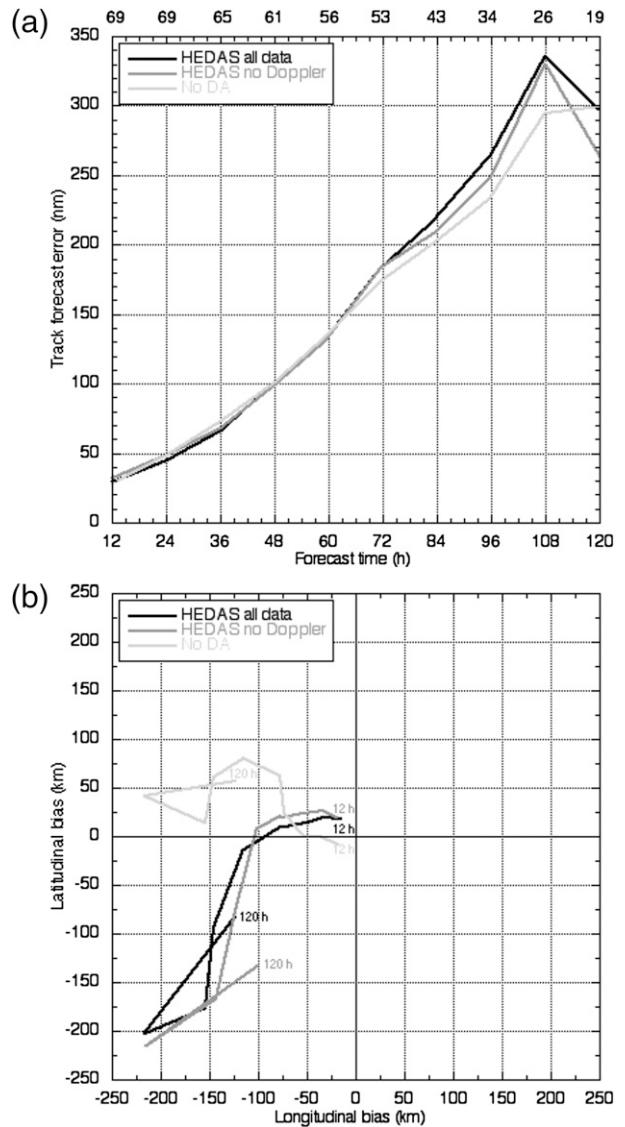
FIG. 1. Homogeneous track forecast (a) errors and (b) biases for the three sets of model runs. The number of cases at each forecast time is presented across the top.

errors is expected, since track is mainly controlled by the flow outside the TC core and nearly all of the observations are from that region. Some large differences in the short range can occur due to initial vortex placement and vortex oscillations that may be predictable within that time. The control forecast bias is notably different from that in the forecasts with data assimilation. The control vortex is placed at the initial location specified by National Hurricane Center (NHC) analyses and comprises any asymmetry representing the tropical cyclone initial motion, whereas no such specification is made for the HEDAS analyses. Since the early-period track forecasts are generally a result of advection of the
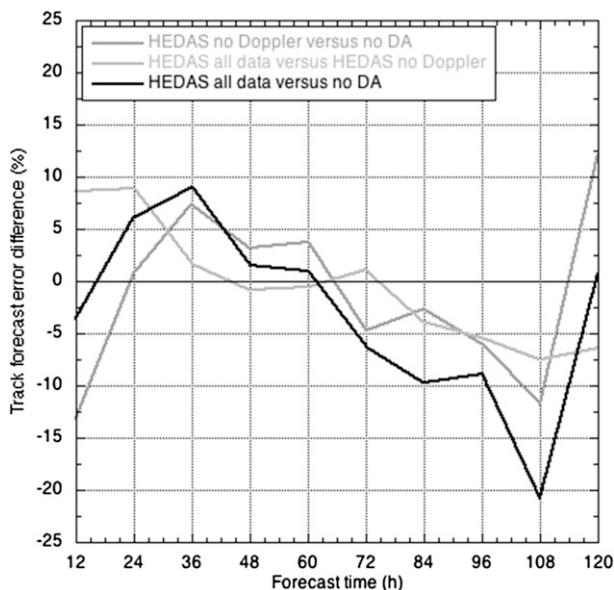
FIG. 2. Pairwise homogeneous differences between possible two-model combinations of the three-model versions from Fig. 1.

vortex, the 12-h forecast biases (Fig. 1b) set up subsequent biases in the forecasts.

The three model versions produce forecasts in which the TC dissipates at different forecast times, thus preventing standard verification at those times. Pairwise comparisons between the three versions allow for sample sizes of up to seven more cases at some forecast times than the homogeneous three-model sample (Fig. 2).[1] The assimilation of aircraft data, both with and without the Doppler radar observations, improves track forecasts by up to 10% versus the no-assimilation control at 24–60 h; improvement is also seen in the small sample at 120 h, though none of the differences are statistically significant (Aberson and DeMaria 1994). The Doppler radar data improve the track forecasts over those with only the standard aircraft observations by nearly 10% during the first 24 h, and the differences are statistically significant at the 95% level; up to 64% of these forecasts are improved by the assimilation of the Doppler radar data. This may be due to better analysis of the outer wind asymmetry steering the vortex by the assimilation of the relatively extensive Doppler radar data versus what is possible with the limited flight-level and dropwindsonde data available in the no-Doppler runs (Fig. 1b). The difference between the two runs becomes small after 24 h.

The relatively large 12-h track forecast errors in the two versions with aircraft data assimilated are likely due to a slight initial vortex displacement by HEDAS: the control is designed so that the vortex initialization puts the center at the assigned location; HEDAS uses observations to analyze the TC core, but Doppler radar data are usually not available at the center location due to the lack of scatterers there, and no explicit location data are assimilated. Though the covariances from the ensemble should alleviate some of the problem caused by the lack of observations near the TC center, the necessarily suboptimal ensemble prevents accurate initial center placement. A technique to assimilate data in a storm-relative framework (Aksoy 2013) alleviates this problem.

b. Intensity

1) STANDARD VERIFICATION

The standard intensity verification (mean and absolute mean differences between the forecast and best track values of maximum wind speed at 10-m altitude) for the same homogeneous set of runs as for track is shown in Fig. 3. Forecast errors remain about constant through 108 h and increase by 120 h, probably due to the small sample size (Fig. 3a). The difference between the forecast intensity errors is expected to be larger than for track errors, since intensity is at least partially controlled by processes within the TC core. The bias (Fig. 3b) increases in time (from negative to positive for all three models). The average biases of the runs with the aircraft data assimilated are small after 48 h and are generally smaller than the average bias of the control runs. The runs with no data assimilated have a substantial high intensity bias after 36 h. The low intensity biases in the short range for all three sets of runs may be due to the vortex spindown issue discussed in Vukicevic et al. (2013), and suggest that this issue is also important in the vortex initialization scheme. This suggests an avenue for further research to improve the representation of the secondary circulation through assimilation and initialization schemes. The particularly large negative bias for the runs without the Doppler radar data may be due to having data only at the aircraft locations (versus the three-dimensional picture that the Doppler radar can provide), which cannot provide an accurate estimate of the intensity. The degradation from the removal of the Doppler data disappears by 36 h into the forecast, after which the results of the two sets of runs with data assimilation mirror each other.

Figure 4 shows the intensity forecast error differences between the three pairs of model runs. As in Fig. 2, the individual comparisons are homogeneous, but the three are not homogeneous with each other. The assimilation of all the aircraft data improves the intensity forecasts by up to 23% compared to the no assimilation runs, and the

---

[1] The differences shown in Figs. 2 and 4 are calculated using difference = 1 − (e1/e2), where e1 and e2 are the errors from the first and second forecast, respectively.
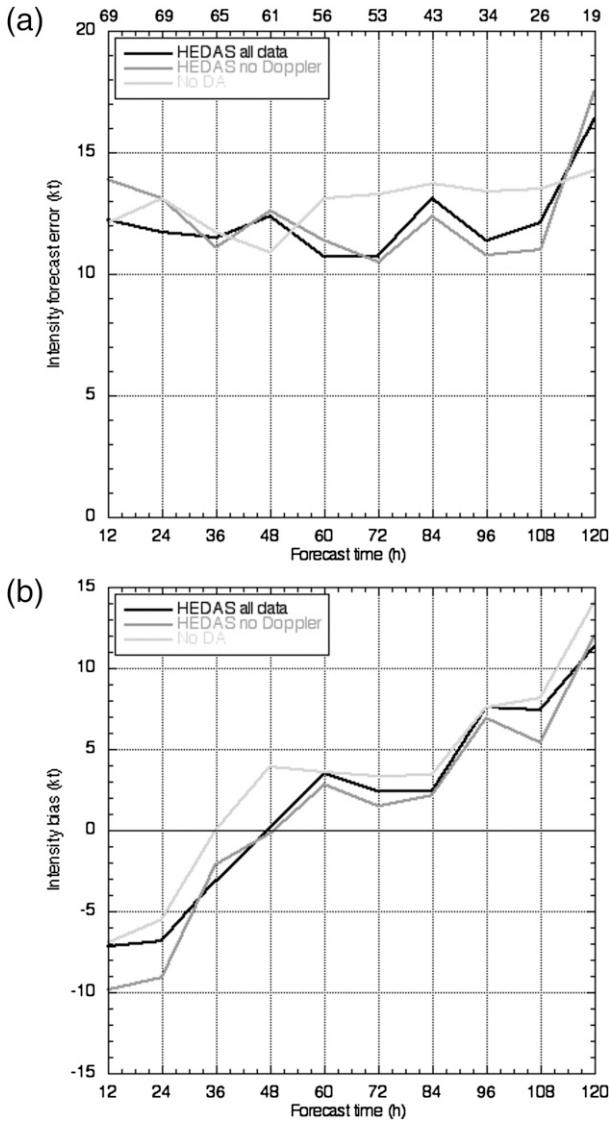
FIG. 3. As in Fig. 1, but for intensity.

FIG. 4. As in Fig. 2, but for intensity.

difference at 24, 36, and 60 h is statistically significant at the 95% level. The Doppler radar data alone improve the forecasts by up to 11% (statistically significant) during the first 24 h of the forecasts. This suggests that most of the improvement at early forecast times is due to the assimilation of the Doppler radar data; the impact (memory) of the Doppler radar data decreases with forecast time, but improvements after 36 h could be attributed to that in the short range. It is unclear why the Doppler radar data degrade the forecasts after 72 h, but the difference is not statistically significant.

2) ALTERNATIVE INTENSITY VERIFICATION

Forecasts are not included using the standard verification technique if any forecast calls for TC dissipation (or if the TC is no longer tracked because it has reached the model boundary) or if the real TC dissipates. To include these forecasts in the verification, Aberson (2008) defined an alternative technique in which these forecasts are included by categorizing all forecasts in a contingency table. All intensity forecasts values are rounded to the nearest 5 kt ($1 \, \mathrm{kt} = 0.5144 \, \mathrm{m \, s^{-1}}$), the same precision as the best track. Contingency tables (not shown) are filled with the count of each forecast–verification pair at each forecast time. The first row and column represent the number of times the TC does, or is forecast to, dissipate, respectively. Each subsequent row and column represents a 5-kt intensity increase. Perfect forecasts are along the diagonal. A skill score,

$$S = (C - E)/(T - E),$$

where C is the number of (correct) forecasts in which the forecast and verification are in the same bin, T is the total number of forecasts, and E is the number of forecasts expected to be correct, is calculated for each contingency table (Panofsky and Brier 1958). The expected number of correct forecasts by chance is

$$E = \Sigma(R_i C_j)/T_i,$$

where R_i and C_j are the numbers of cases in the ith row and jth column, respectively. The skill score is unity if all cases are correctly predicted ($T = C$), and less than or equal to zero for no skill.

The three sets of runs show considerable skill at all forecast times except at 84 h in that with the Doppler

radar data assimilated (Fig. 5a). Differences between the standard and alternative verifications are evident. The runs with all data assimilated are better than the control through 108 h in the standard verification; they are better than the control at all forecast times except 36, 48, 84, and 96 h in the alternative technique. The runs with the Doppler radar data assimilated are better than those with just the regular aircraft data at 12, 24, 48, 60, and 120 h in the standard verification, but are better than those with just the regular aircraft data at 24, 60, 72, and 108 h in the alternative technique. This suggests that a large number of cases had forecasts of dissipation at different lead times in the three sets of model runs.

Figure 5b shows the proportion of cases for which the categorical forecasts are good or poor (within three or more than six categories from the verification, respectively). The assimilation of all aircraft data increases the proportion of cases with good forecasts versus the control at all forecast times. Interestingly, the assimilation of the Doppler radar data also increases the proportion of forecasts that are poor at all forecast times from 36 to 108 h compared to the set of forecasts with the standard aircraft observations assimilated, an effect that is not accounted for in the standard verification. However, the set with the Doppler radar data assimilated has fewer poor forecasts than the control.

3) RAPID INTENSITY CHANGE

An important issue in intensity forecasting is the ability of the models to predict large changes in intensity during short time periods. For this study, rapid intensification is defined as an increase in intensity of 30 kt or more during a 24-h period; rapid weakening is defined as a decrease in intensity of at least 25 kt during a 24-h period. To remove cases in which rapid weakening occurs due to land interactions, all cases in which the forecast or the observed TC moved over land are removed from the sample. To account for the best track having a 5-kt granularity and the model forecasts having one of 1 kt, the forecasts are binned as in the alternative intensity verification above. Thus, if the best track or forecast intensity increases by six bins or decreases by five bins, then the criteria for rapid intensity changes are met. If a 24-h period during which both the model forecast and the best track have rapid intensity changes overlap in time, then the forecast is said to be correct; since multiple 24-h periods of one forecast can overlap a single instance of the other, both are counted. If the model predicts a rapid intensity change episode that does not overlap with one in the best track, this is counted as a false alarm; if the model fails to predict rapid intensity change during the entire period in which it existed in the best track, this is considered a missed
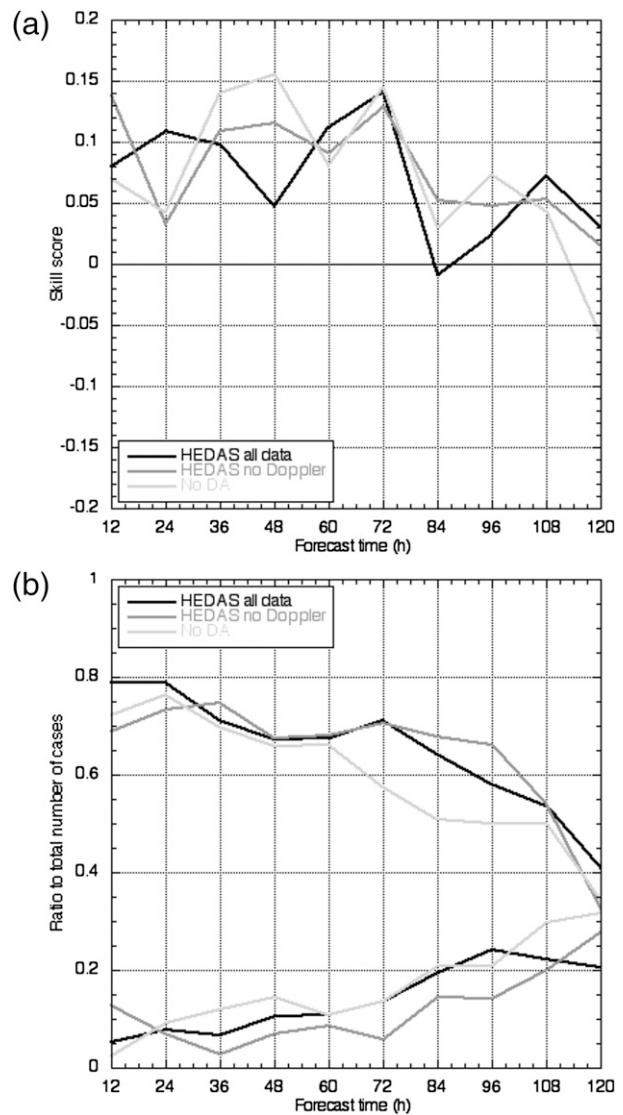
FIG. 5. (a) As in Fig. 1a, but for the skill score from the alternative intensity verification technique, and (b) the proportion of cases with forecasts within three categories of the verification (upper set of lines) and more than six categories from the verification (lower set of lines) from the alternative intensity verification technique.

forecast. The fourth possibility (neither the model forecast nor the best track have a rapid intensity change period) is not counted, as they comprise the vast majority of cases and are uninteresting.

The rapid intensification and rapid weakening forecast verifications are summarized in a categorical performance diagram (Roebber 2009; Fig. 6). The probability of detection is plotted against the success ratio; the bias is represented by the lines emanating from the origin, with the diagonal having no bias; the threat score or critical success index is represented by the curved lines. The
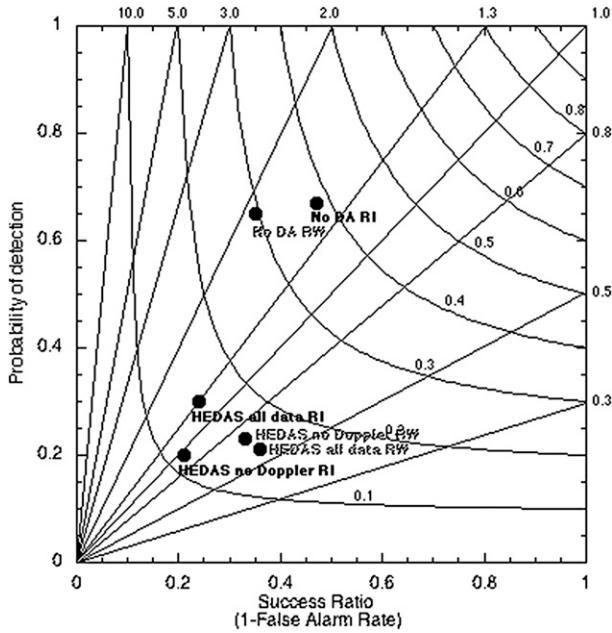
FIG. 6. Categorical performance diagram for all three versions of the model and for rapid intensification (RI) and rapid weakening (RW).

success ratio and probability of detection would both be unity (upper-right-hand corner of the figure) if the forecasts were perfect (every rapid intensity change event correctly forecast and every forecast event verifying). A total of 30 rapid intensification episodes occurred in the best track during the 120-h forecasts. The control forecast had 51 rapid intensification episodes, whereas the runs with and without the Doppler radar data assimilated forecast had 17 and 19 such periods, respectively. The control is therefore about 3 times more likely to predict rapid intensification than the data assimilation runs and thus is more likely to correctly forecast rapid intensification episodes and to have false alarms. Of the 30 observed episodes, the control correctly forecast 20 events, whereas the runs with and without the Doppler radar data correctly forecast 9 and 6 events, respectively; the remainder of the cases are thus missed forecasts. A total of 24 of the control forecasts verified, whereas both sets of HEDAS runs had 4 forecasts verify correctly. The control had 27 false alarms, compared with 13 and 15 false alarms for the runs with and without the Doppler radar data assimilated, respectively.

A total of 61 rapid weakening episodes occurred within 120 h of the initialization times. The control forecast rapid weakening events 52 times; the runs with and without the Doppler radar data assimilated forecast it 16 and 15 times, respectively. The control is 4 times more likely to forecast rapid weakening than the others,

again suggesting a higher likelihood of correct forecasts and false alarms. Of the 52 rapid weakening events forecast by the control, 34 events were correctly forecast, and the runs with and without the Doppler radar data assimilated correctly forecast 11 and 12 events, respectively. The control had 34 false alarms, and the two sets of runs with data assimilated had 9 and 10 false alarms, respectively. Because the rapid intensity change episodes are spread evenly throughout the forecasts in time in the control and data assimilation runs, the reason that the control has far more rapid intensity change episodes than the runs with the data assimilated is unknown.

c. Structure

Recent cases in which weak, but very large TCs made landfall leading to catastrophic destruction (Hurricane Ike in Texas, Hurricane Sandy in the U.S. Northeast) show the importance of verifying structure forecasts in addition to track and intensity. NHC provides TC structure forecasts of maximum gale-, storm-, and hurricane-force surface (10 m) wind speed radii in four quadrants and radius of maximum wind speed in addition to the standard track and intensity forecasts, and best track estimates of these quantities except the radius of maximum wind speed. Insufficient surface observations make operational and best track wind speed radii estimates uncertain; a typical 30 n mi (~55.6 km) hurricane-force wind speed radius could be in error by 50% or more (Rappaport et al. 2009). Because of these difficulties, structure forecasts are not generally verified against the best track. Despite this uncertainty, these values can be combined into one metric, the integrated kinetic energy (IKE; Powell and Reinhold 2007), that helps to alleviate these problems.

To construct a reasonable wind field from the forecast parameters available in the Automated Tropical Cyclone Forecast system from which to approximate IKE the following simple assumptions are made:

1) Gale-, storm-, and hurricane-force winds all extend outward from 70% of R_{MW}, and this radius is denoted r. Because the volume represented between r and R_{MW} is small, changes to this percentage do not significantly change the final IKE calculation.
2) The wind speed at R_{MW} is identically V_{max}. This greatly simplifies the calculations, and is not a particularly poor assumption (Vukicevic et al. 2014).
3) The wind speed varies linearly between each wind speed radius, allowing for simple mean wind speed calculations in different annuli ($\overline{V}_{1,2}$); $\overline{V}_{1,2}$ changes slightly for realistic ratios of the outer to the inner radius, but can be approximated for simplicity as

$$\overline{V}_{1,2} = 0.625(V_1 + V_2).$$

4) The wind field can be represented approximately as a vortex of wavenumbers 0 and 1, with the latter having an amplitude of about 15 kt, or near the average forward motion of TCs. The volume (with a depth of 1 m) of a particular quadrant wind speed between two thresholds outward from R_{MW} is thus approximated as

$$A = \pi[R_2^2 - 1/4(R_1 + R_2)^2],$$

where R_2 is the larger of the two radii. Inside R_{MW}, the volume of a quadrant is

$$A = 0.51\pi R_{MW}^2.$$

The mass is calculated as the surface air density ($\rho = 1.15\,\mathrm{kg\,m^{-3}}$) multiplied by the volume. If a particular radius does not exist, its value is set to 0. To calculate IKE in a particular quadrant, the radii that exist (between 2 and 5 radii in each quadrant) are ordered from smallest to largest in each quadrant, and the sum of IKE between each consecutive one is calculated:

$$\mathrm{IKE} = \frac{1}{2}\rho\Sigma A_{i,j}(\overline{V}_{i,j})^2,$$

where i and j represent consecutively ordered radii.

Model IKE forecasts are verified against the best track estimates (Fig. 7). Because the model postprocessor that calculates wind speed radii was developed after the control run model fields were lost, no comparison with these runs is possible. The forecasts with the Doppler radar data provide up to 20% improvement in IKE forecasts through 36 h versus those with the standard aircraft data, probably due to improved initial analysis from the large amount of data available, but mostly degrades forecasts thereafter. However, the difference between the forecast errors is not statistically significant. The biases grow linearly with forecast time, and are always positive, meaning that both sets of runs tend to predict larger and/or stronger storms than are suggested by the best track.

4. Forecast examples

Two cases in which the aircraft data provided particularly large and consistent track and intensity forecast impacts are discussed, even though these differences were not always improvements. In the first case, the assimilation of aircraft data greatly improved the track forecast of Hurricane Irene while degrading the intensity forecasts. In the second, the assimilation greatly improved both the track and intensity forecasts for Tropical Storm Tomas even though none of the forecasts were highly accurate. Though
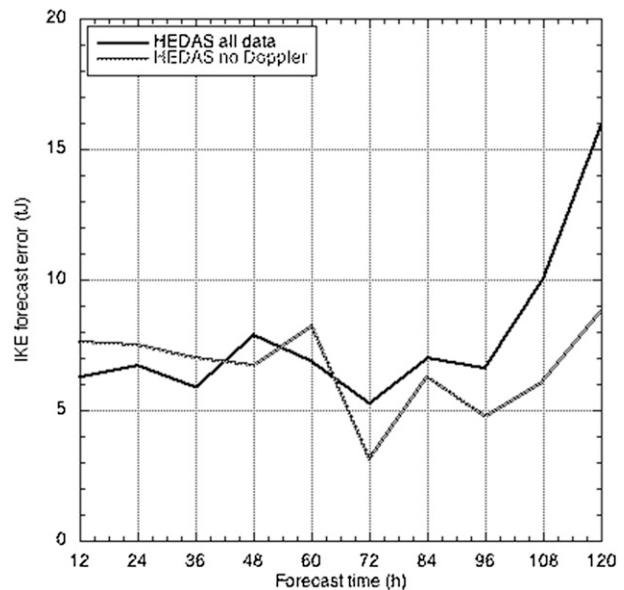
FIG. 7. As in Fig. 1a, but for IKE.

only location and intensity are generally verified for these models, the fact that one forecast parameter can be consistently improved and another degraded in the same model run suggests a more holistic approach to hurricane model verification is needed.

a. Hurricane Irene: 1200 UTC 26 August 2011

On 26 August 2011, Hurricane Irene was approaching the North Carolina coastline as a category 2 hurricane (Fig. 8). The no-assimilation control run forecast Irene to move more slowly and less eastward than the two versions with aircraft data assimilated. Both those forecasts had very small cross-track errors, and the assimilation of the Doppler data allowed for large improvements in the along-track errors throughout the forecast. Despite the improved track forecasts, the two runs with aircraft data assimilated forecast Irene to become much stronger than the no-assimilation control, thus degrading the intensity forecasts. Both versions of the model with data assimilated predicted rapid weakening between 24 and 48 h; Irene weakened 20 kt during this period, just below the threshold for rapid weakening. This results in a false alarm forecast for rapid weakening. The IKE forecasts were not appreciably different from each other.

Figure 9 shows initial condition wind fields at three levels for the no-assimilation control and the run with all the aircraft data assimilated; the third run is not appreciably different from the other HEDAS run in this case, and is not shown. The control (left) has higher wind speeds than the HEDAS run at all levels, but the major asymmetry is consistent at all levels in the HEDAS run, whereas it rotates clockwise with height in the control.
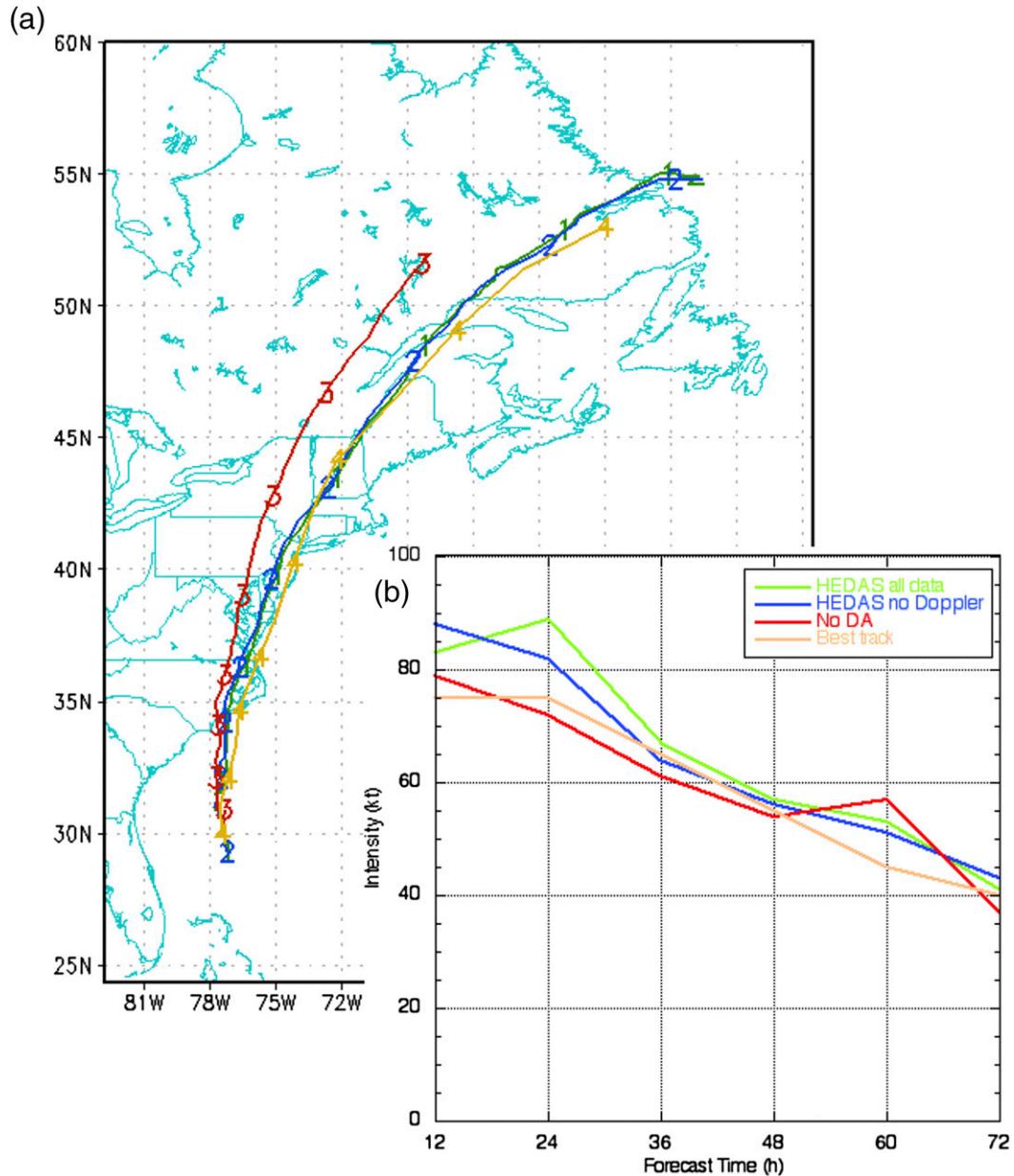
FIG. 8. (a) Track and (b) intensity forecasts for Hurricane Irene initialized at 1200 UTC 26 Aug 2011. For track, symbols are plotted every 12 h in the forecast. In (a), the green line is HEDAS all data, the blue line is HEDAS no Doppler, the red line is the No DA control, and the yellow line is the best track. The full track until dissipation is shown in (a); intensity forecasts in (b) are shown through 72 h.

By 24 h, the midlevel maximum wind speeds had rotated counterclockwise to the northeast quadrant in the control, but remained steady in the HEDAS run. The synoptic conditions in the runs (all based on the operational GFS) were nearly identical and are not shown. The differences in the wind speed asymmetries between the runs account for the more westerly track (and improved) forecast by the no-assimilation control versus the HEDAS runs, but the relatively high wind speeds at all levels in these runs accounted for the degraded intensity forecasts.

b. Tropical Storm Tomas: 1200 UTC 4 November 2010

Tropical Storm Tomas was located south of Jamaica and had turned northeastward toward Haiti at
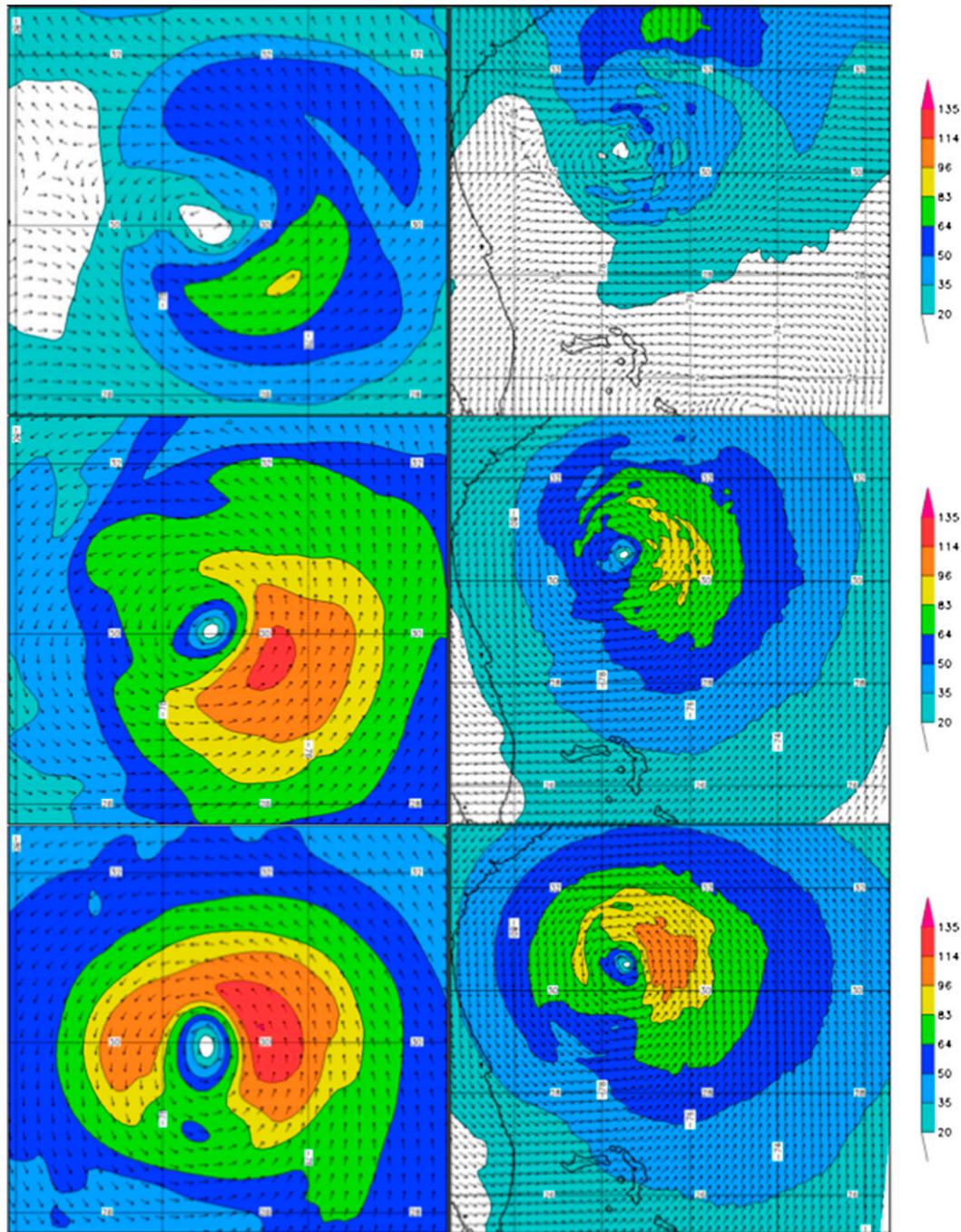
FIG. 9. Wind velocity initial conditions (kt) (inner mesh only) for Hurricane Irene initialized at 1200 UTC 26 Aug 2011, showing the (left) no-assimilation control and (right) all-data HEDAS at (top) 200, (middle) 500, and (bottom) 850 hPa on the innermost mesh of each run.

1200 UTC 4 November 2010. The assimilation of air-craft data slightly degraded the across-track forecasts (Fig. 10), but the along-track errors were greatly re-duced leading to large forecast track improvements. None of the forecasts correctly predicted the in-teraction between Tomas and an upper-level low that

turned the storm toward the east; all three forecasts predicted dissipation too early due to the too-rapid predicted northward motion. The assimilation also allowed for substantial improvement to intensity forecasts, though all three model versions predicted premature dissipation: the version with no Doppler
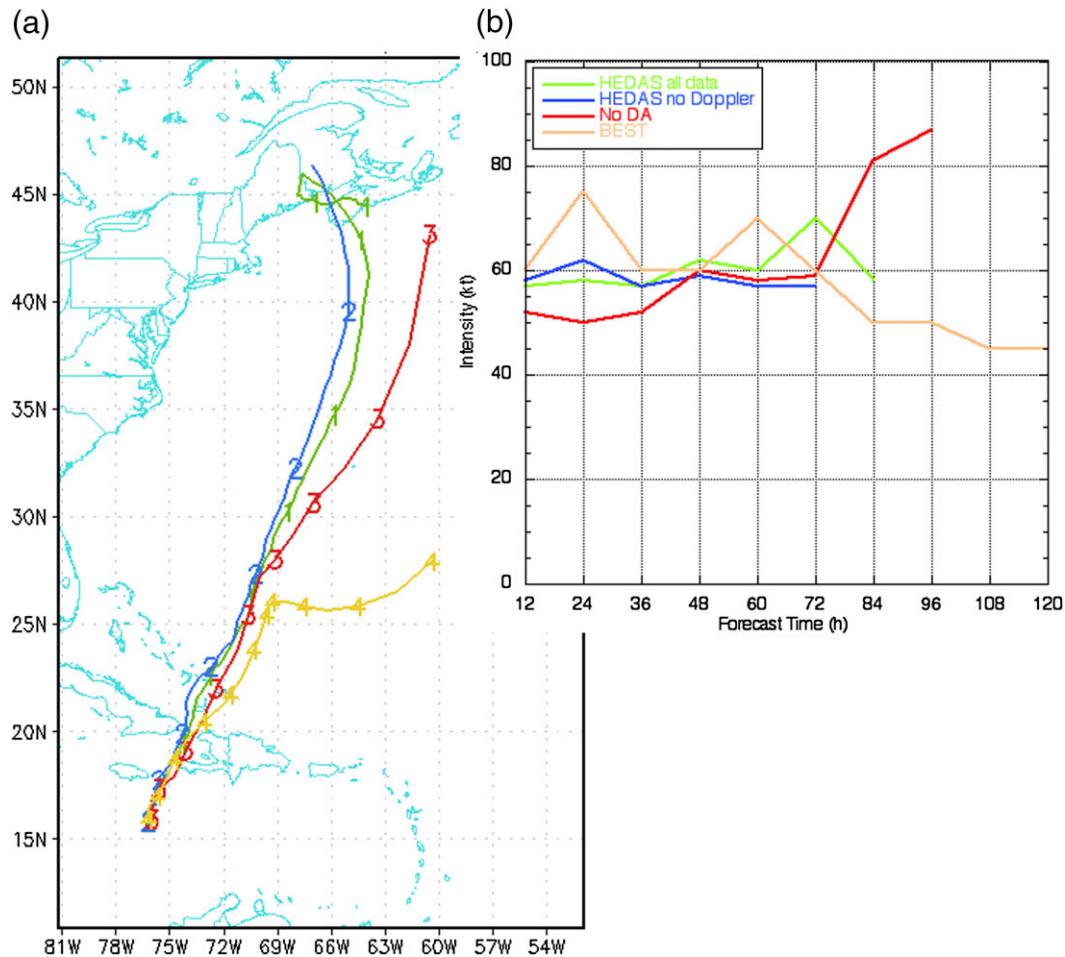
FIG. 10. As in Fig. 8, but for Tropical Storm Tomas initialized at 1200 UTC 4 Nov 2010.

radar data assimilated predicted dissipation at 72 h when the best track intensity was 60 kt; the version with all data assimilated and the control predicted dissipation at 84 h and 96 h, respectively, when the best track intensity was 50 kt. All three correctly forecast that rapid intensity change would not happen. The IKE forecasts were not appreciably different from each other in this case.

The no-assimilation control was initialized with an idealized vortex initialization (Fig. 11) that was weaker than those for the HEDAS runs. This allowed Tomas in the no-assimilation control to be steered northward by the relatively slow low-layer flow, whereas the stronger TCs in the HEDAS runs were steered by a faster, deeper-layer flow. Though the HEDAS TCs were initially stronger than that in no-assimilation control, the faster motion caused the TC to move over cooler water and into a high-shear environment, causing weakening, whereas the slower-moving TC stayed over warm water and was able to rapidly intensify toward the end of the forecast partially due to baroclinic

forcing. The improved initial condition improved both the track and intensity forecasts in this case.

5. Conclusions and discussion

The impact of high-resolution aircraft observations on forecasts is investigated with an EnKF-based data assimilation system, HEDAS, and HWRF. A total of 83 cases from 20 TCs during 2008–11 are considered. HEDAS assimilates available observations from NOAA P3s and G-IV, and Air Force C-130s in 1-h cycles. Observation types assimilated include airborne Doppler radar wind superobservations, temperature and wind velocity from aircraft flight-level measurements and dropwindsondes, and Stepped-Frequency Microwave Radiometer 10-m wind speed retrievals.

The important results are as follows:

1) For track, the assimilation of aircraft data improves forecasts by up to 10% at 24–60 h; large improvement is also seen in the small sample at 120 h. The Doppler
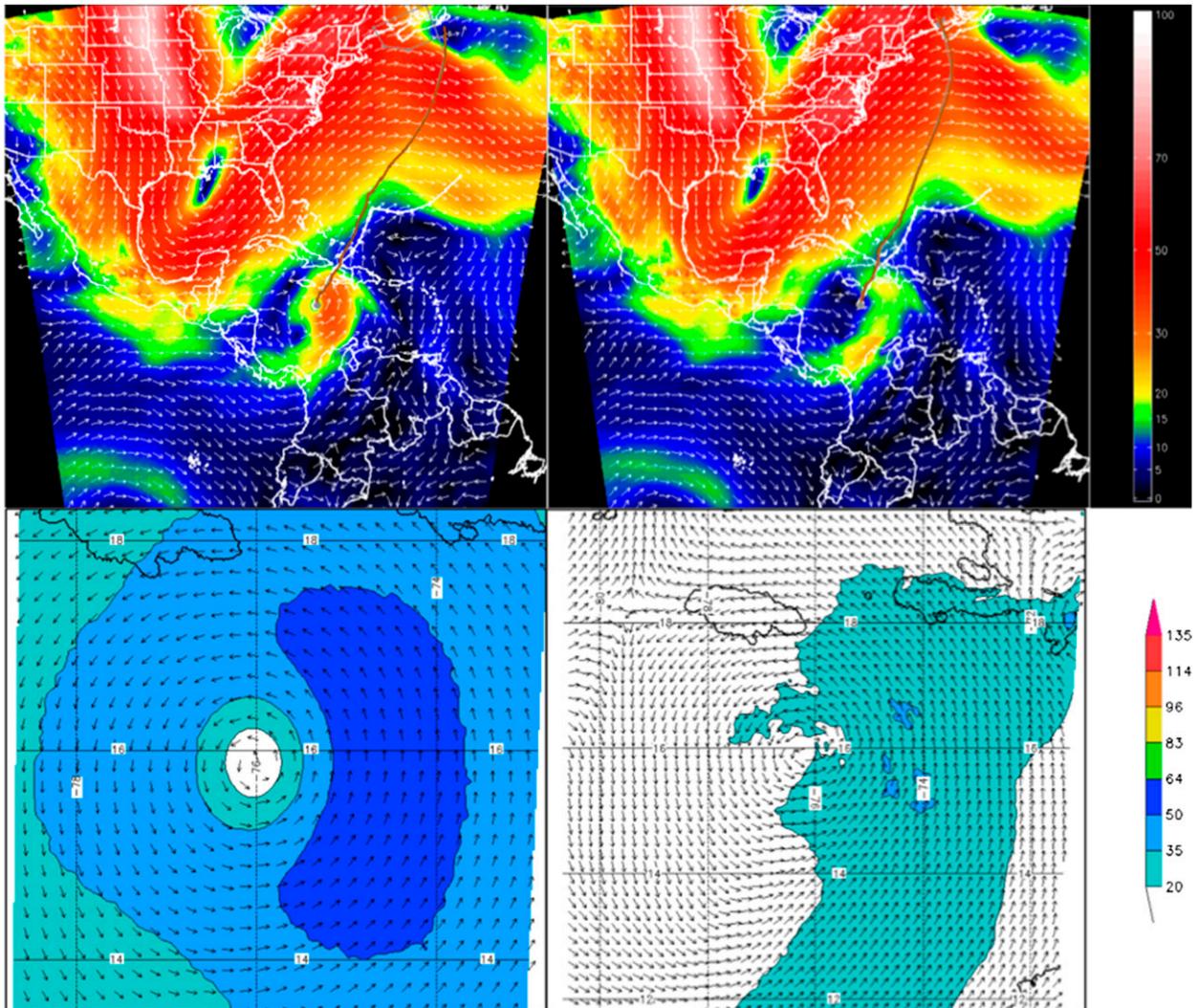
FIG. 11. Wind velocity initial conditions for Tropical Storm Tomas initialized at 1200 UTC 4 Nov 2010 for the (left) no-assimilation control and (right) all-data HEDAS for the (top) large-scale 850–200-hPa deep-layer mean and (bottom) 700 hPa on the innermost mesh.

radar data alone improve the forecasts by up to 10% during the first 24 h, and the difference is statistically significant at the 95% level; up to 64% of these forecasts are improved by the assimilation of the Doppler radar data.

2) For intensity, the assimilation of all the aircraft data improves the forecasts by up to 23% compared to the no-assimilation runs, and the difference at 24, 36, and 60 h is statistically significant at the 95% level. The Doppler radar data alone improve the forecasts by up to 11% (statistically significant) during the first 24 h of the forecasts.

3) For structure, the forecasts with the Doppler radar data provide up to 20% improvement in IKE forecasts through 36 h, though the difference between the forecast errors is not statistically significant.

Recent upgrades to the quality-control software for the airborne Doppler radar are able to provide more superobservations in the boundary layer than previously available (J. Gamache 2014, personal communication). Recent 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). And HEDAS was enhanced to, among other things, update the vertical acceleration of the wind to improve initial analyses of the vertical component of the velocity, all in a storm-relative framework (Aksoy 2013). These upgrades were implemented for the 2012 hurricane season, and results will be presented separately.

REFERENCES

Aberson, S. D., 2008: An alternative tropical cyclone intensity forecast verification technique. *Wea. Forecasting,* **23,** 1304–1310, doi:10.1175/2008WAF2222123.1.

——, 2010: 10 years of hurricane synoptic surveillance (1997–2006). *Mon. Wea. Rev.,* **138,** 1536–1549, doi:10.1175/2009MWR3090.1.

——, and M. DeMaria, 1994: Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Wea. Rev.,* **122,** 2804–2815, doi:10.1175/1520-0493(1994)122<2804:VOANBH>2.0.CO;2.

——, 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, doi:10.1175/BAMS-87-8-1039.

——, J. Cione, C.-C. Wu, M. M. Bell, J. Halverson, C. Fogarty, and M. Weissmann, 2010: Aircraft observations of tropical cyclones. *Global Perspectives on Tropical Cyclones: From Science to Mitigation,* J. C. L. Chan and J. D. Kepert, Eds., Vol. 4, World Scientific, 227–242.

Aksoy, A., 2013: Storm-relative observations in tropical cyclone data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.,* **141,** 506–522, doi:10.1175/MWR-D-12-00094.1.

——, 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, doi:10.1175/MWR-D-11-00212.1.

——, S. D. Aberson, T. Vukicevic, K. J. Sellwood, S. Lorsolo, and X. Zhang, 2013: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD's HEDAS: Evaluation of the 2008–11 vortex-scale analyses. *Mon. Wea. Rev.,* **141,** 1842–1865, doi:10.1175/MWR-D-12-00194.1.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99,** 10 143–10 162, doi:10.1029/94JC00572.

Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.,* **94,** 329–343, doi:10.1175/BAMS-D-12-00071.1.

Gopalakrishnan, S. G., 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, doi:10.1175/WAF-D-11-00055.1.

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, doi:10.1175/MWR-D-11-00079.1.

——, ——, 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, doi:10.1175/2010MWR3456.1.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Kurihara, Y., M. A. Bender, and R. J. Ross, 1993: An initialization scheme of hurricane models by vortex specification. *Mon. Wea. Rev.,* **121,** 2030–2045, doi:10.1175/1520-0493(1993)121<2030:AISOHM>2.0.CO;2.

——, ——, R. E. Tuleya, and R. J. Ross, 1995: Improvements in the GFDL Hurricane Prediction System. *Mon. Wea. Rev.,* **123,** 2791–2801, doi:10.1175/1520-0493(1995)123<2791:IITGHP>2.0.CO;2.

Panofsky, H. A., and G. W. Brier, 1958: *Some Applications of Statistics to Meteorology.* The Pennsylvania State University, 224 pp.

Powell, M. D., and T. A. Reinhold, 2007: Tropical cyclone destructive potential by integrated kinetic energy. *Bull. Amer. Meteor. Soc.,* **88,** 513–526, doi:10.1175/BAMS-88-4-513.

Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting,* **24,** 395–419, doi:10.1175/2008WAF2222128.1.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting,* **24,** 601–608, doi:10.1175/2008WAF2222159.1.

Vukicevic, T., A. Aksoy, P. Reasor, S. D. Aberson, K. J. Sellwood, and F. Marks, 2013: Joint impact of forecast tendency and state error biases in ensemble Kalman filter data assimilation of inner-core tropical cyclone observations. *Mon. Wea. Rev.,* **141,** 2992–3006, doi:10.1175/MWR-D-12-00211.1.

——, E. Uhlhorn, P. Reasor, and B. Klotz, 2014: A novel multiscale intensity metric for evaluation of tropical cyclone intensity forecasts. *J. Atmos. Sci.,* **71,** 1292–1304, doi:10.1175/JAS-D-13-0153.1.

Whitaker, J., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Yeh, K.-S., X. Zhang, S. G. Gopalakrishnan, S. Aberson, R. Rogers, F. D. Marks, and R. Atlas, 2012: Performance of the experimental HWRF in the 2008 hurricane season. *Nat. Hazards,* **63,** 1439–1449, doi:10.1007/s11069-011-9787-7.

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.,* **38,** L15810, doi:10.1029/2011GL048469.

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