Evaluation of Tropical Cyclone Structure Forecasts in a High-Resolution Version of the Multiscale GFDL fvGFS Model

ANDREW T. HAZELTON

Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey

LUCAS HARRIS AND SHIAN-JIANN LIN

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

(Manuscript received 21 September 2017, in final form 16 January 2018)

ABSTRACT

A nested version of the cubed-sphere finite-volume dynamical core (FV3) with GFS physics (fvGFS) is capable of tropical cyclone (TC) prediction across multiple space and time scales, from subseasonal prediction to high-resolution structure and intensity forecasting. Here, a version of fvGFS with 2-km resolution covering most of the North Atlantic is evaluated for its ability to simulate TC track, intensity, and finescale structure. TC structure is evaluated through a comparison of forecasts with three-dimensional Doppler radar from P-3 flights by NOAA's Hurricane Research Division (HRD), and the structural metrics evaluated include the 2-km radius of maximum wind (RMW), slope of the RMW, depth of the TC vortex, and horizontal vortex decay rate. Seven TCs from the 2010–16 seasons are evaluated, including 10 separate model runs and 38 individual flights. The model had some success in producing rapid intensification (RI) forecasts for Earl, Edouard, and Matthew. The fvGFS model successfully predicts RMWs in the 25–50-km range but tends to have a small bias at very large radii and a large bias at very small radii. The wind peak also tends to be somewhat too sharp, and the vortex depth occasionally has a high bias, especially for storms that are observed to be shallow. Composite radial wind shows that the boundary layer tends to be too deep, although the outflow structure aloft is relatively consistent with observations. These results highlight the utility of the structural evaluation of TC forecasts and also show the promise of fvGFS for forecasting TCs.

1. Introduction and prior work

Although tropical cyclone (TC) track and intensity forecasts continue to improve (albeit at a slower rate for intensity), TC structure forecasts can be inconsistent with observations. As numerical models improve in their horizontal grid resolution, evaluation of TC structure forecasts is becoming a major research focus. TC structure is important to forecasts of how TC hazards vary spatially and also to assessments of model forecast skill—whether, for example, an intensity forecast is accurate "for the right reasons." The metrics for evaluating TC track and intensity are well known and widely used, and the datasets used in this evaluation are also well established [National Hurricane Center (NHC) Hurricane Database (HURDAT2); Landsea et al. (2015)]. However, there are fewer established techniques for verifying and evaluating TC structure, and detailed analyses of structure forecasts are important for making further improvements to numerical models.

A few studies have attempted to verify horizontal and vertical structure forecasts of tropical cyclones. For example, Houze et al. (2007) showed that an MM5 simulation of Hurricane Rita (2005) at 1.67-km horizontal resolution was able to successfully predict its secondary eyewall formation (e.g., Willoughby et al. 1982). Vertical and horizontal metrics based on TC kinematic structure have been used to evaluate TC simulations. Fierro et al. (2009) studied the impact of horizontal resolution on Weather Research and Forecasting (WRF) Model forecasts of Hurricane Rita and found that higher resolution led to a smaller radius of maximum winds (RMW) with a more upright eyewall. This RMW–slope relationship was consistent with several observational studies (Stern and Nolan 2009; Stern et al. 2014; Hazelton and Hart 2013; Hazelton et al. 2015). Zhang et al. (2012) used observational composites of

Corresponding author: Andrew Hazelton, andrew.hazelton@noaa.gov

DOI: 10.1175/WAF-D-17-0140.1

boundary layer structure based on dropsonde data from aircraft to improve the boundary and surface-layer structure of the Hurricane Weather Research and Forecasting (HWRF) Model. Building on that study, Zhang et al. (2015) further evaluated the impact of improvements to the boundary layer parameterization of the HWRF Model on simulated TC structure, using both dropsonde and airborne radar data. Zhang et al. (2017) expanded on this work even more, showing that smaller vertical diffusion in the boundary layer led to stronger inflow in a shallower boundary layer, which increased both convection in the inner-core region and also inward transport of angular momentum, making storms more likely to undergo rapid intensification (RI). Similarly, Gopalakrishnan et al. (2013) studied the impact of vertical diffusion on boundary layer structure and TC intensity in the HWRF Model. That study found that reducing vertical diffusion produced a shallower PBL that was more realistic based on aircraft observations, as well as a smaller TC eyewall and stronger TCs. Nolan et al. (2009a,b) used observations (Black et al. 2007) from dropsondes in Hurricane Isabel (2003) to validate, compare, and improve boundary layer parameterizations in the WRF Model. These studies also found improvements in simulations of TC intensity and structure by increasing the model resolution from 4 to 1.33 km. Marks et al. (2016) used airborne Doppler radar observations to evaluate HWRF initialization and forecast RMW and found that the model suffers from a high bias (too large of an RMW) in the initial conditions. While most prior studies focus on kinematic structure and verification, Zick and Matyas (2016) analyzed precipitation structure in TCs in the Gulf of Mexico by developing and analyzing structural shape metrics. That study speculated that such a technique could be used to verify numerical model predictions of TC structure.

Those studies indicated that inner-core measurements are a critical component of verifying structure forecasts of TCs, and the development of structural metrics (both composite and at individual times) is critical to allow for a detailed comparison between the model and observations. In this study, we will analyze structural metrics based on known important features of TC structure to analyze the forecasts of a new, developmental high-resolution nested model. This will demonstrate the utility of structure analysis in TC forecast verification and also highlight the strengths and weaknesses of the current version of the cubed-sphere finite-volume dynamical core (FV3) with Global Forecast System (GFS) physics (fvGFS). This comparison with observations will motivate improvements to the model physics in future studies to better represent TC structure (and intensity) and also provide further guidance for the evaluation of both vertical and horizontal structures in TC models.

2. Data

a. Model configuration

The model used for the forecasts in this study uses the NOAA/Geophysical Fluid Dynamics Laboratory (GFDL) FV3 dynamical core (Lin and Rood 1997; Lin 1997, 2004) with similar physical parameterizations as those used in the National Centers for Environmental Prediction (NCEP) GFS. The combined FV3 dynamical core with (modified) GFS physics is a global model that can be regionally refined through stretching and nesting, as described below. The model is initialized using the global GFS analyses interpolated to the fvGFS global and nested grids. The radiation scheme is the Rapid Radiative Transfer Model for GCMs (RRTMG; Iacono et al. 2008). The boundary layer scheme is based on Han and Pan (2011). For deep and shallow convection, the model employs the simplified Arakawa–Schubert (SAS) scheme (Arakawa and Schubert 1974, Han and Pan 2011). The GFS microphysics scheme (based on Zhao and Carr 1997) has been replaced for this study by a 6-class single-moment microphysics scheme developed at GFDL. It is briefly described in Chen and Lin (2013) and is similar to the 6-class scheme of Lin et al. (1983). Chen et al. (2018, unpublished manuscript) and Bender (2017) examined the skill of the 13-km global fvGFS on TC track, intensity, and genesis forecasts and found that the model had improved skill in TC prediction compared with the operational GFS model, especially when using the new 6-class microphysics scheme. Test cases of a slightly lower-resolution version of the nested fvGFS used in this study also showed improved intensity forecasts from the new microphysics scheme compared to the current GFS scheme.

For this study, the global fvGFS grid is a c1152 grid (~8.5-km resolution) stretched by a factor of 1.5 over the North Atlantic basin, using the transformation of Schmidt (1977; see also Harris et al. 2016), to achieve grid-cell widths of about 5.7 km over the North Atlantic. A factor-of-3 static nested grid (Fig. 1a) with two-way interaction (Harris and Lin 2013) is then placed to reach ~2-km grid-cell widths over the western North Atlantic. For these experiments, the model was not coupled to an ocean model, which may contribute to a positive intensity bias at later forecast times (shown later). This capability is in development for future studies.

b. Radar data

This study uses NOAA Lockheed WP-3D Orion (P-3) tail-mounted Doppler radar analyses. The three-dimensional

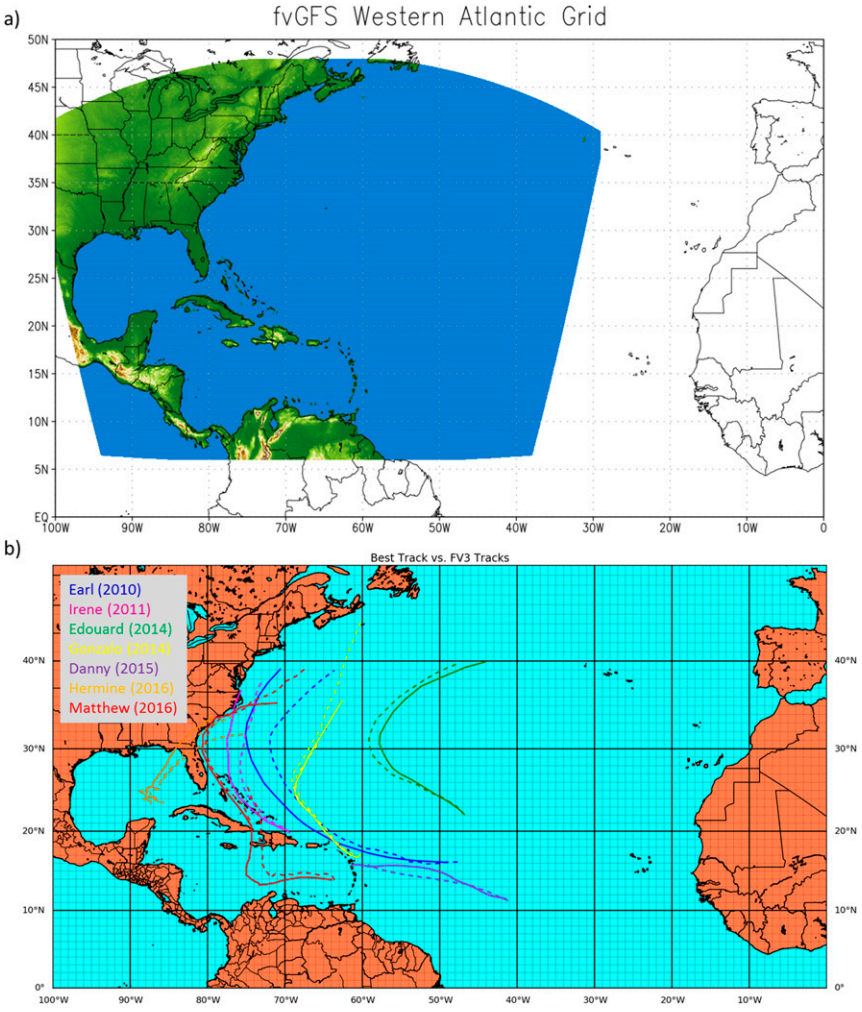

FIG. 1. (a) The nested Atlantic domain used in the forecasts. (b) Tracks of seven Atlantic TCs used in this study (solid lines) and the corresponding fvGFS forecasts (dashed lines).

wind is derived from the raw radar data using the method of Gamache et al. (2004). This technique has been found to produce RMS errors of less than $\sim 3\,\mathrm{m\,s^{-1}}$ for tangential wind, although vertical velocity is more difficult to capture with accuracy (Rogers et al. 2012; Lorsolo et al. 2013). The radar data are obtained in a three-dimensional swath along each leg of a flight as it passes through and around a TC. The swaths usually give good coverage of the eye and eyewall, especially if the inner core is small enough. However, to get better coverage of both the eyewall and outer rainbands, a commonly used technique is to average several flight legs to form a "merged analysis." This technique does not work well for vertical velocity, as it tends to smooth out strong updrafts and downdrafts (which are more transient than the horizontal vortex). However, the merged analyses have been demonstrated to be effective for horizontal wind, which tends to evolve on a slower time scale than the typical 3–4-h time on station (Rogers et al. 2013).

c. *Cases analyzed*

This study is based on 10 forecasts of seven different Atlantic TCs from 2010 to 2016. Figure 1b shows the tracks of these TCs (during the forecast period), as well as the corresponding model tracks. Table 1 lists the TCs analyzed, as well as the number of P-3 flights for each TC and the model initialization times and initial TC intensities. Four of the TCs had one forecast, and three had two forecasts. All forecasts were initialized after TC genesis had already occurred, but Fig. 1b shows that the forecasts cover a wide range of locations and tracks in the basin, from westward movement at low latitudes to recurvature in the midlatitude westerlies. All TCs selected had radar data from the NOAA Hurricane

TABLE 1. List of the TCs used in this study, as well as the number of P-3 flights included in the dataset and the initialization times of the fvGFS.

Storm	Year	No. of P-3 flights	Initialization time(s)	Initial intensity (kt)	Run length (h)
Earl	2010	10	0000 UTC 28 Aug 2010	45	168
Irene	2011	7	0000 UTC 23 Aug 2011	80	120
Edouard	2014	6	1200 UTC 13 Sep 2014	45	120
Gonzalo	2014	3	0000 UTC 13 Oct 2014	45	120
			0000 UTC 15 Oct 2014	115	84
Danny	2015	3	1200 UTC 19 Aug 2015	45	120
Hermine	2016	3	0000 UTC 30 Aug 2016	30	96
			0000 UTC 31 Aug 2016	30	72
Matthew	2016	6	0000 UTC 29 Sep 2016	55	168
			0000 UTC 5 Oct 2016	110	120

Research Division (HRD) flights. Most cases were started early in the TC life cycle, since the model was initialized from the relatively coarse GFS analyses and was not able to initialize the inner-core structure fully (a vortex initialization algorithm for fvGFS is a subject of ongoing work), while the second Matthew and Gonzalo cases were initialized later to cover a period with more observational data.

3. Combined results: Track and intensity verification, and structure evaluation

a. Track and intensity verification

When evaluating model TC forecasts, it is standard to compare the track and intensity forecasts to HURDAT2 (Landsea et al. 2015). The track and intensity errors (Figs. 2a,b) are compared with the forecasts from statistical models, a common baseline for forecast skill. For track, this is the climatology and persistence (CLIPER) model (Aberson 1998), and for intensity the Statistical Hurricane Intensity Forecast model (SHIFOR) baseline (Knaff et al. 2003) is used. The NHC official forecast errors for these cases are also shown. The track forecast is skillful at all forecast hours (Fig. 2a), with errors slightly higher than NHC at hours 12–96 but lower at hour 120. The intensity forecasts show skill after a 24-h spinup and have lower errors than NHC from hours 48 to 120. Figure 2c shows the intensity bias. There is a low bias at shorter lead times, likely due to the spinup problem as well as the underprediction of RI in some cases. However, at longer lead times the intensity bias is positive, with many of the storms too strong as they recurve over the North Atlantic. The bias at longer leads may be due to the fact that the current version of fvGFS is not coupled to an ocean model and therefore does not account for changing SSTs due to wind-induced mixing. A planned model upgrade will include a simplified ocean model to account for this effect. Also, based on a comparison of one to two test cases with the cumulus scheme turned off (not shown), the cumulus parameterization may also be partially responsible for some of the positive bias.

b. Individual structure metrics

1) COMPARISON OF FVGFS FORECAST AND RADAR-DERIVED METRICS

The model is evaluated beyond the track and intensity errors by examining structural metrics and comparing the model forecast structure with the observed structure from the P-3 radar data. The first metric analyzed is the azimuthal-mean RMW at $z = 2$ km. A metric designed to show the horizontal decay rate of the tangential winds, known as the modified Rankine exponent (here called α), was defined according to Mallen et al. (2005) by

$$\frac{V_1}{V_2} = \left(\frac{R_2}{R_1}\right)^{\alpha}.$$

In this study, R_1 is the 2-km RMW described above, and R_2 is $3 \times \mathrm{RMW}_{2\mathrm{km}}$. The tangential wind speeds at R_1 and R_2 are V_1 and V_2, respectively. Thus, the α parameter measures how "sharp" the azimuthal-mean wind peak is.

Two metrics of vertical structure are analyzed. The first is the vortex depth, defined as the height at which the tangential wind decays from the value at $z = 2$ km to 75% of this 2-km value. For stronger storms (2-km tangential wind $> 50\,\mathrm{m\,s^{-1}}$), this value was relaxed to 50%, as the 75% value was found to give artificially low vortex depths (perhaps because of more scatterers aloft in stronger storms). This metric of vortex depth did not specifically account for the tilt of the vortex, although vortex tilt would be useful to examine as a separate metric in further analyses. The second measure of the vertical structure analyzed was the slope of the eyewall

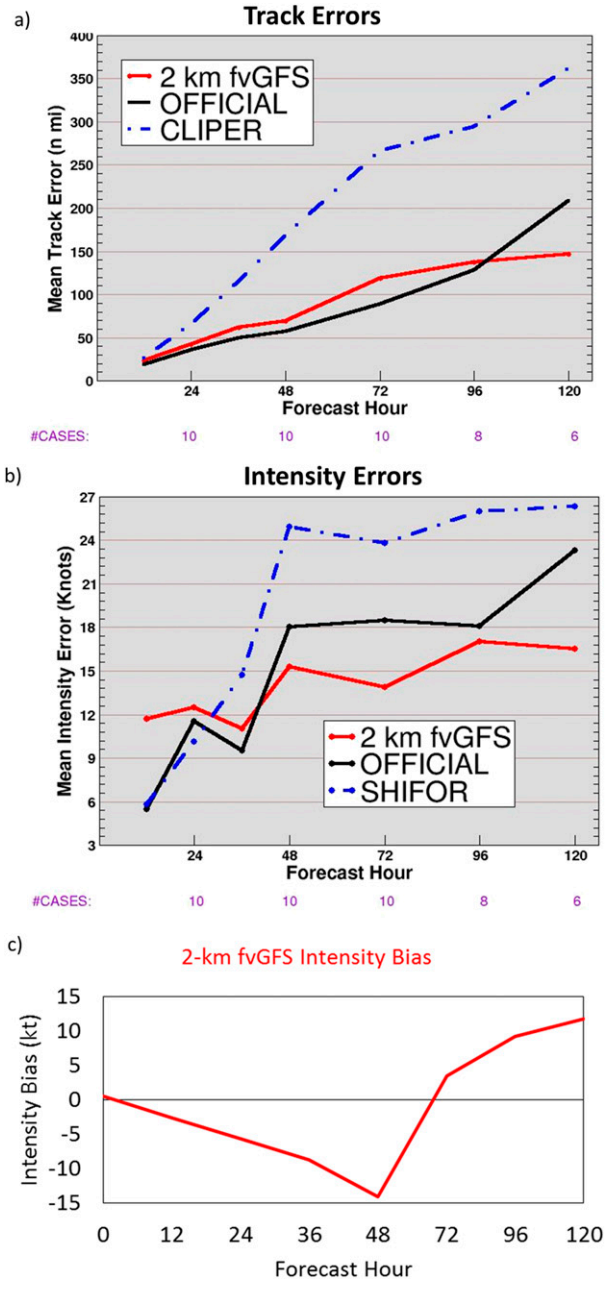

FIG. 2. (a) The fvGFS track errors (red), CLIPER track errors (blue), and NHC official track errors (black) (n mi; 1 n mi = 1.852 km). (b) The fvGFS intensity errors (red), SHIFOR intensity errors (blue), and NHC official intensity errors (black) (kt). (c) The fvGFS intensity bias (kt).

(e.g., Stern and Nolan 2009; Stern et al. 2014; Hazelton and Hart 2013; Hazelton et al. 2015). In this study, the azimuthal-mean RMW between $z = 2$ and 8 km was used to calculate the eyewall slope. Figure 3 shows scatterplots of each structure metric, with the radar-derived observations shown along the x axis and the model forecasts along the y axis. Due to the low sample size at any given forecast hour, all forecast hours are included together. The shading is proportional to the intensity forecast bias.

The RMW results (Fig. 3a) show that the model forecasts match best with observations in the 25–50-km range. For small observed radii, the model is generally too large (although there are two cases where the RMW is smaller than 25 km), and the model RMW also has a tendency to be too small when the observed RMW is very large. This is likely connected to the fact that the storms were often too intense at later forecast hours (as most of the cases where the RMW is too small have a positive intensity bias) and may also indicate that the model does not always correctly forecast RMW increases due to secondary eyewalls (e.g., Willoughby et al. 1982). It is also worth noting that the RMW range appears to be much smaller in the model than in observations, indicating that the model tends to drift toward a preferred range of values (based on resolution and physics) and does not capture the extreme contraction or rapid expansion seen in some cases.

The results for the vortex decay parameter α (Fig. 3b) show that the model has a tendency to have too sharp of a wind peak. This could also be connected to the high-intensity bias, as Willoughby (1990) showed a similar structure with a broader wind peak when the maximum wind was lower, although such a relationship is not clear from Fig. 3b. Both the intensity and vortex decay biases could be tied to the lack of ocean coupling. In addition, they could also be related to the structure of secondary eyewalls, which tend to lead to a broader wind field and weaker wind maximum. It is also worth noting that many of the cases with very large observed RMW were not included in this sample, as the $3 \times$ RMW radius used to define the vortex decay parameter extended outside the range of the radar data in these cases.

For vortex depth (Fig. 3c), the bias is less consistent. For relatively shallow observed storms, there is a wide range of model values, although the overall tendency is for the model vortices to be too deep. This is again consistent with the overall high bias in intensity, as most of the cases that were too deep were also too strong. The model did a relatively good job of capturing the observed vortex depth for very deep storms (above 10 km or so).

RMW slope (Fig. 3d) shows a similar pattern as the RMW itself. Many of the cases with model RMWs that are too small also have too upright of an RMW (smaller slope). This is consistent with the eyewall size and slope relationship found by Stern and Nolan (2009), Hazelton and Hart (2013), and Stern et al. (2014). However, there does not appear to be a clear connection between

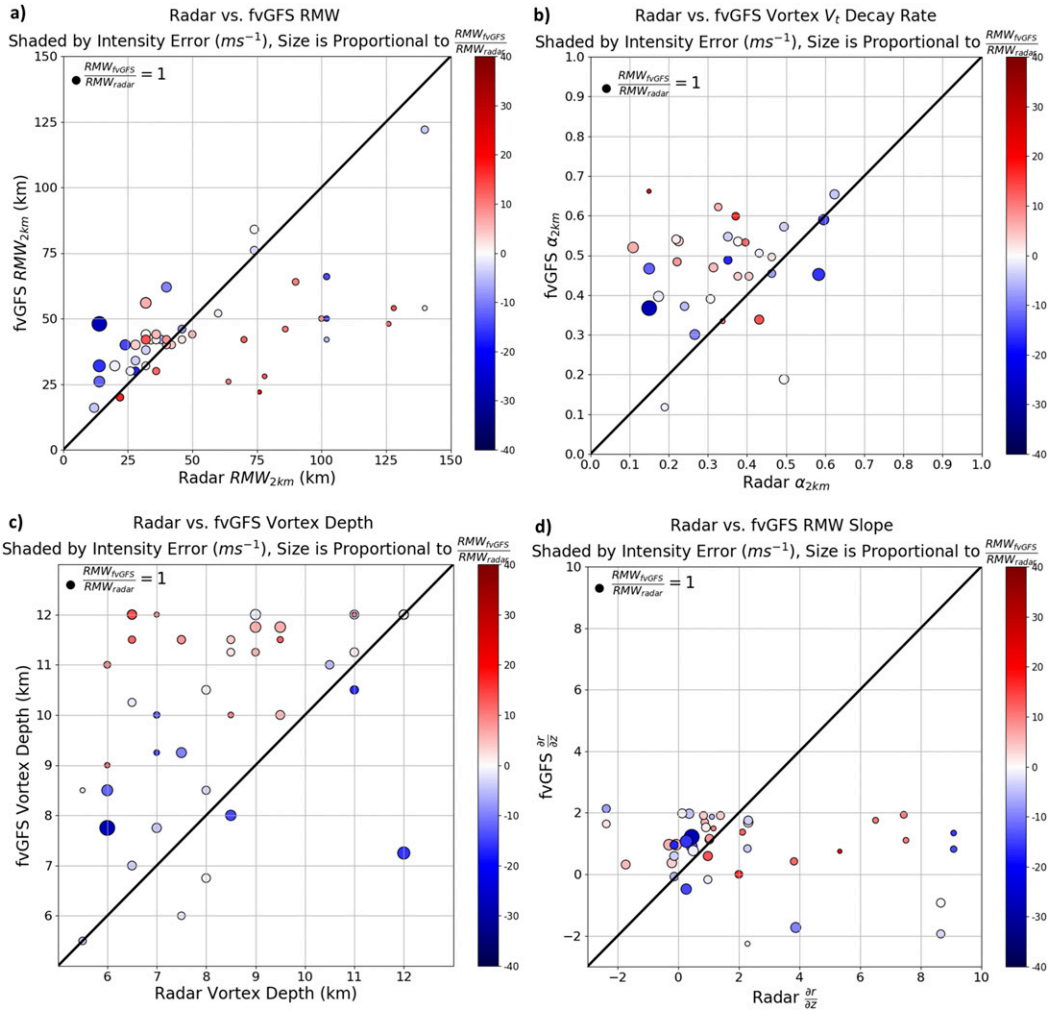

FIG. 3. (a) Radar-derived RMW at $z = 2$ km vs fvGFS forecast RMW at $z = 2$ km. The circles are shaded based on the intensity error (m s^{-1}; positive errors are overforecasts). The size of the circle is proportional to the ratio of the model RMW to the observed RMW (scale shown in the top left of the panel). (b) As in (a), but for the vortex decay parameter α. (c) As in (a), but for the vortex depth. (d) As in (a), but for the slope of the RMW.

intensity errors and slope biases. The largest model RMW slope is about 2 (or 60° away from the vertical), while the observed values have a larger range. The model also has fewer cases with slopes near 0. There seems to be a ''preferred range'' of slopes in the model data between 0.5 and 2 that is not reflected in the observations. This could be due to model resolution, as well as the use of a convective scheme, in which parameterized updrafts are upright rather than sloped.

2) COMPARISON OF ERRORS OF DIFFERENT METRICS

Table 2 shows relevant statistics for the structure forecast errors for each metric. Table 2 lists the correlation coefficient between the radar-derived metric and the fvGFS forecast. Also listed are the mean and median of the percentage errors as well as the ''normalized'' percentage errors (calculated by dividing the percentage errors by the maximum percentage error for each metric). This normalized percentage error is intended to help account for variations in the metrics, that is, the fact that the alpha parameter is constrained between 0 and 1, the RMW slope can be both positive and negative, and the RMW and vortex depth are constrained to be positive.

The results are mixed (in terms of which structure parameters are ''best'' forecast). The RMW forecasts had the highest correlation with the observations but were relatively similar in percentage error and normalized percentage error to vortex depth and α. Vortex depth also had a relatively high correlation with the

TABLE 2. Correlation between the fvGFS forecast structure metrics and the corresponding radar-observed values, as well as the median percentage errors and the median normalized percentage errors for each metric.

Metric	Model–radar correlation	Median model % error (0%–100%)	Median model normalized % error (0–1)
RMW	0.56	32.3	0.13
α	0.18	38.9	0.1
Vortex depth	0.49	30.9	0.37
RMW slope	0.29	96.8	0.06

observations, but the normalized percentage error was the highest of all structure metrics. The RMW slope had a very high percentage error, but as noted above, this error metric is different because of the fact that it can be negative. When accounting for this fact with the normalized errors, the RMW slope actually was the best-forecast metric. Finally, α had the lowest correlation with the observations, but had the second-lowest normalized percentage error. Thus, the best-forecast parameter varies significantly based on the error metric chosen. Another important caveat is the fact that several of the cases had an undefined α due to $3 \times$ RMW being outside the range of the radar data. Further complicating matters is the relationship between the different metrics [see section 3b(3)]. Thus, it cannot be stated definitively whether the model is better at forecasting the horizontal or vertical structure, at least based on this sample.

3) RELATIONSHIPS BETWEEN STRUCTURE METRICS IN THE MODEL AND OBSERVATIONS

To further explore the structure forecasts via the individual structural metrics, the relationship between metrics is explored in both the model and observations. The results are listed in Table 3. The first relationship explored is that between RMW and RMW slope. Stern and Nolan (2009) and Stern et al. (2014) found a relationship between RMW size and slope, and that result is confirmed in the observations here ($r = 0.41$, $p < 0.01$). However, the relationship is virtually nonexistent in the model. This is in contrast to the HWRF Model, which did reproduce this relationship relatively well (Zhang et al. 2015). Based on the findings of that paper, it is possible that some of the structure forecast issues in this study are due to the poor parameterization of vertical diffusion in the boundary layer. For the RMW–α relationship, the model has a relatively strong relationship, with smaller RMWs being associated with a sharper wind peak. This relationship is much weaker in the observations, and the observational findings are consistent with the work of Mallen et al. (2005), who found no relation between RMW and α. As will be shown later in one of the case studies, the model is likely underrepresenting the magnitude of the observed secondary wind maxima, which weakens the relationship between the vortex decay rate and the inner RMW in the observed set of cases. Paired with the stronger peak winds often observed, it appears that the model does not always spread out angular momentum as much as is seen in observed TCs.

There is an observed relationship between vortex depth and slope, with deeper vortices when the eyewall is more upright. This is consistent with a study of the vertical decay rate of the tangential winds in both observed and simulated TCs by Stern and Nolan (2011). This relationship is much less evident in the model forecasts. Finally, the model shows a relatively strong correlation between α and RMW slope. This is less evident in the observations (not statistically significant) and is likely tied to the fact that the RMW–α relationship is also stronger in the model.

These structural relationship comparisons show that the model is able to reproduce some of the observed variability, particularly the relationship between RMW and vortex depth. However, the model misses some of the key observed connections, especially those between the RMW and RMW slope. The structure composites

TABLE 3. Relationships between various structure metrics in both the observational radar dataset and the model dataset. Only model times with radar data available are included.

Relationship	Observed r (p value)	Model r (p value)
RMW–RMW slope	0.41 (<0.01)	−0.04 (0.77)
RMW–vortex depth	−0.29 (0.05)	−0.24 (0.11)
RMW–α	−0.23 (0.21)	−0.46 (<0.01)
Vortex depth–RMW slope	−0.39 (<0.01)	−0.16 (0.29)
Vortex depth–α	0.02 (0.93)	−0.11 (0.56)
RMW slope–α	−0.25 (0.18)	−0.47 (<0.01)

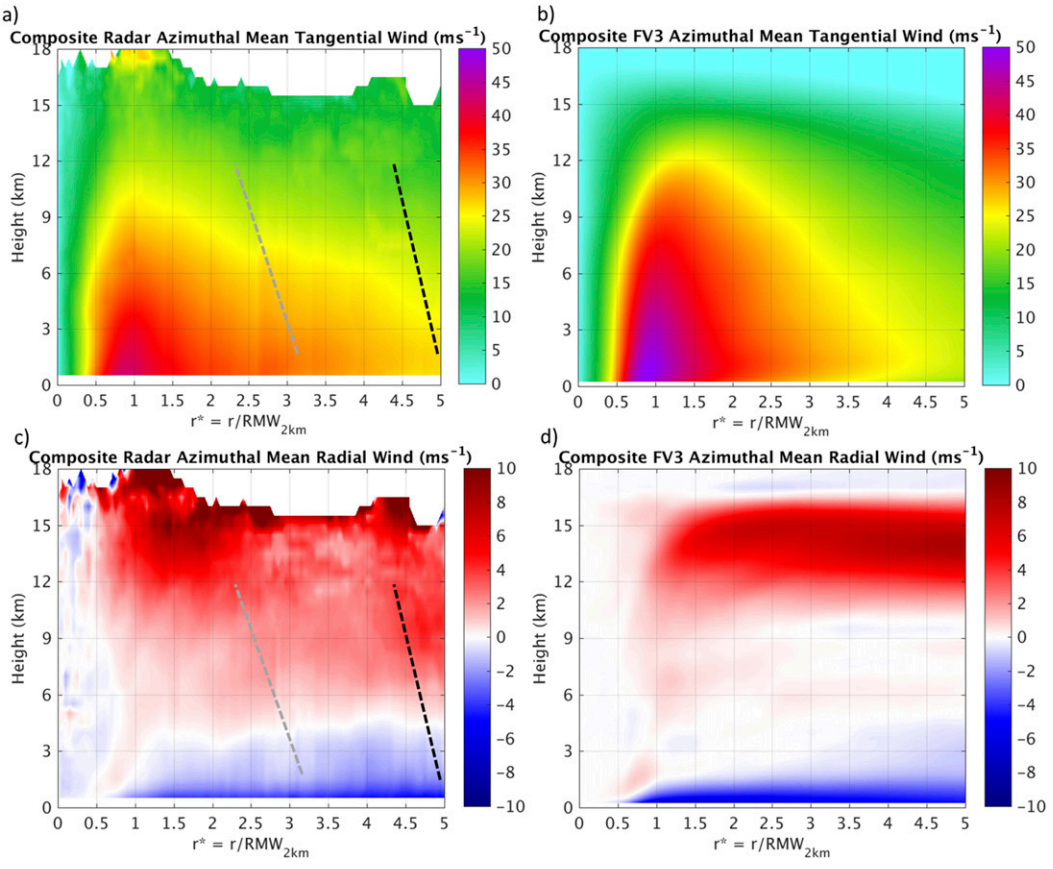

FIG. 4. (a) Composite tangential wind (m s^{-1}) from the tail-mounted Doppler radar from 38 flights used in this study. The radial coordinate is normalized by the RMW at $z = 2$ km. The dashed gray line shows the approximate normalized radii at which the composite contains data from 75% of the cases, and the dashed black line shows the approximate normalized radii at which the composite contains data from 50% of the cases. (b) Composite tangential wind (m s^{-1}) from fvGFS forecasts matching the radar times used to create the composite in (a). Once again, the radial coordinate is normalized by the RMW at $z = 2$ km. (c) As in (a), but for the radial wind (m s^{-1}). (d) As in (b), but for the radial wind (m s^{-1}).

(examined next) will build on these findings, and case studies will show more details of exactly what is occurring in some of the forecasts.

c. *Model versus radar composites*

To analyze the structure beyond discrete metrics, composites of radar structure versus model structure were next examined. For a homogeneous comparison, only model times matching the radar time (at a 6-hourly resolution) were included in the model composites. For both the model and radar data, the data are normalized by the radius of maximum wind (at $z = 2$ km) before compositing, in order to avoid excessive smoothing caused by variations in storm size. This technique has been used in multiple studies for compositing the P-3 radar data (e.g., Rogers et al. 2013; Reasor et al. 2013; DeHart et al. 2014; Hazelton et al. 2015).

Figure 4 shows azimuthal-mean tangential and radial wind for the radar data and fvGFS forecast composites in the normalized coordinate system. The tangential wind composites are consistent with the intensity and vertical structure verifications. The model composite has stronger tangential wind than the radar composite. In addition, the model composite vortex is too deep, with the 30 m s^{-1} contour extending up to almost 12 km (compared to ~9 km in the observations). The radar composite also shows a broader vortex than the model composite. However, the radar composite is likely less trustworthy at larger normalized radii because of a smaller sample size (see dashed lines in Fig. 4 showing the approximate radii where the composite is missing 25% and 50% of the cases). The radial wind composites show that the strongest boundary layer inflow appears to be too deep in the model, perhaps because of issues with

the boundary layer itself being too deep (e.g., Zhang et al. 2012), which is a known bias of the PBL scheme. In addition, although there is weak outflow in the model composite above 6 km, it is not as strong as in the observations. The outflow in the model is more concentrated at upper levels and is maximized around z = 12–16 km. This peak height appears to be relatively consistent with the observations. The radial location of the outflow is slightly different than the observations, but, once again, sample size may be an issue with the observed data because coverage tends to be limited at these upper levels, particularly for weaker storms where the hydrometeors do not extend as high into the troposphere.

The composite structure was also examined within the context of the vertical shear (Fig. 5). Vertical shear has been shown to have a significant impact on TC vertical and radial velocity structure in previous radar studies (e.g., Reasor et al. 2013; DeHart et al. 2014). Following the methodology used in those (and other) studies, the radial velocity is separated into four shear-relative quadrants: downshear right (DSR), downshear left (DSL), upshear left (USL), and upshear right (USR). For the radar data, the shear used is the 850–200-hPa shear vector from the Statistical Hurricane Intensity Prediction Scheme (SHIPS) archive (e.g., DeMaria and Kaplan 1994). This shear vector is calculated over a 200–600-km annulus radially outward from the center of the TC. For the model data, a similar annulus is created to calculate the 850–200-hPa shear.

The model shear-relative composites of the radial flow are generally similar to the observational composites. The low-level inflow is maximized downshear, with only a very shallow layer of inflow in the upshear region and outflow from ~1 to 4 km. The model also captures the observed upper-level inflow from ~4 to 10 km in the upshear region due to the shear-induced asymmetric secondary circulation that also tends to be associated with sinking motion in the upshear region. The only major difference observed is that the model also has an area of upper-level inflow between 7 and 12 km in the DSL quadrant. In the observed composite, this is a relative minimum in the outflow magnitude, but does not change to actual inflow. Although not shown here (as the observational vertical velocity data are less reliable than horizontal wind), the model composite vertical velocity has a double maximum at low levels and aloft, while the radar data show only the upper-level peak. This is likely tied to the radial flow differences and indicates that the model's response to shear is not perfectly consistent with the observations. However, these composites show that the model is replicating a wavenumber-1 asymmetry generated by vertical shear. Although the individual structure metrics show that the model has areas in which to improve, the wavenumber-1 structure is critical for skillful forecasts of TC intensity and impacts.

4. Individual case studies

Besides the composite results, the track, intensity, and structure forecasts of individual cases are analyzed to provide further insight into model biases and to provide opportunities for a more detailed structural analysis. Once again, the track and intensity forecasts are verified against best-track data, and structural comparisons are made between the model data and the P-3 radar data.

a. Hurricane Earl initialized 0000 UTC 28 August 2010

The first case study is Hurricane Earl (2010). Earl was a long-track Cape Verde hurricane that has been extensively studied by observational (e.g., Rogers et al. 2015) and numerical studies (e.g., Stevenson et al. 2014; Chen and Gopalakrishnan 2015). The fvGFS forecast was initialized at 0000 UTC 28 August 2010, while Earl was a tropical storm over the central Atlantic.

The track and intensity of this Earl forecast versus the best-track data are shown in Fig. 6. The 3-hourly model wind speed has been smoothed with a running 6-hourly mean. Figure 6 also shows the azimuthal-mean RMW (at z = 2 km) from the fvGFS forecasts and also from the P-3 radar data matched to the closest 6-h time. The track plot shows that the model forecast track had a slight right-of-track bias that increased with time, particularly west of 60°W. The intensity plots show that the storm was actually too strong at hour 6 of the forecast during the spinup process, but after a realistic vortex intensity was established around hour 12, the model forecast was very close to reality until hour 72 or so. In fact, the model TC deepened by 35 kt (where 1 kt = 0.51 m s^{-1}) from hours 36 to 60, close to the observed deepening of 40 kt during this period, and above the rapid intensification threshold of 30 kt in 24 h (Kaplan et al. 2010). The intensification of the observed storm halted around hour 60 due to an eyewall replacement cycle (ERC; Cangialosi 2011; Rogers et al. 2015), but the model TC did not stop intensifying until approximately hour 78. The weakening period that began around hour 132 in the model as the storm reached higher latitudes was also seen in the observations, although the observed decay rate was greater. The RMW of the forecast did not get as small as the observed RMW during the RI phase (28 km in the radar data vs 40 km in the model), but was generally close to the observations during the steady period. The RMW was actually too small as the model TC

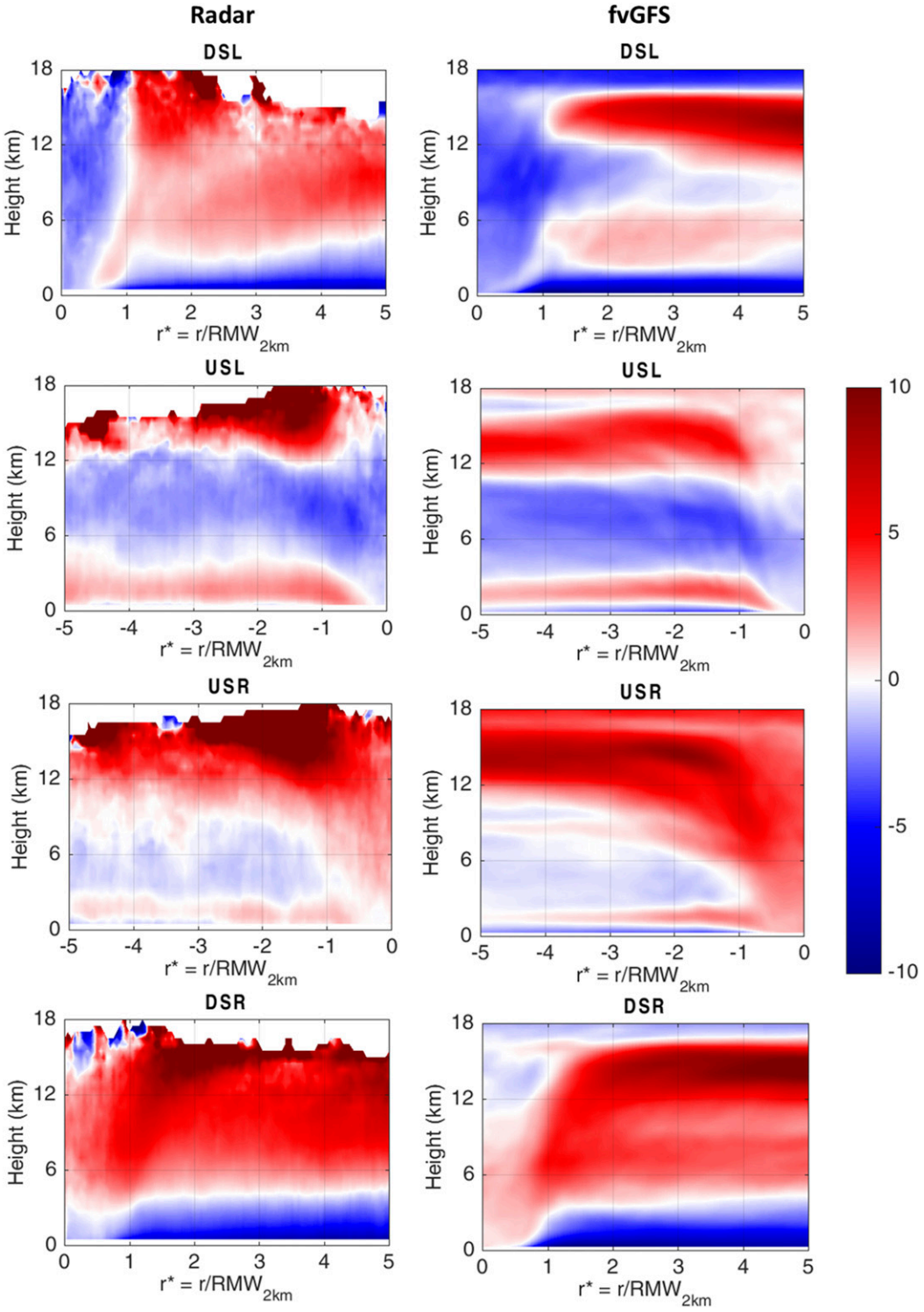

FIG. 5. Composite radial wind (m s^{-1}) in each of four shear-relative quadrants (DSL, USL, USR, and DSR) from (left) the tail-mounted Doppler radar data from the 38 flights used in this study and (right) the corresponding fvGFS forecasts. The radial coordinate is normalized by the RMW at $z = 2$ km.

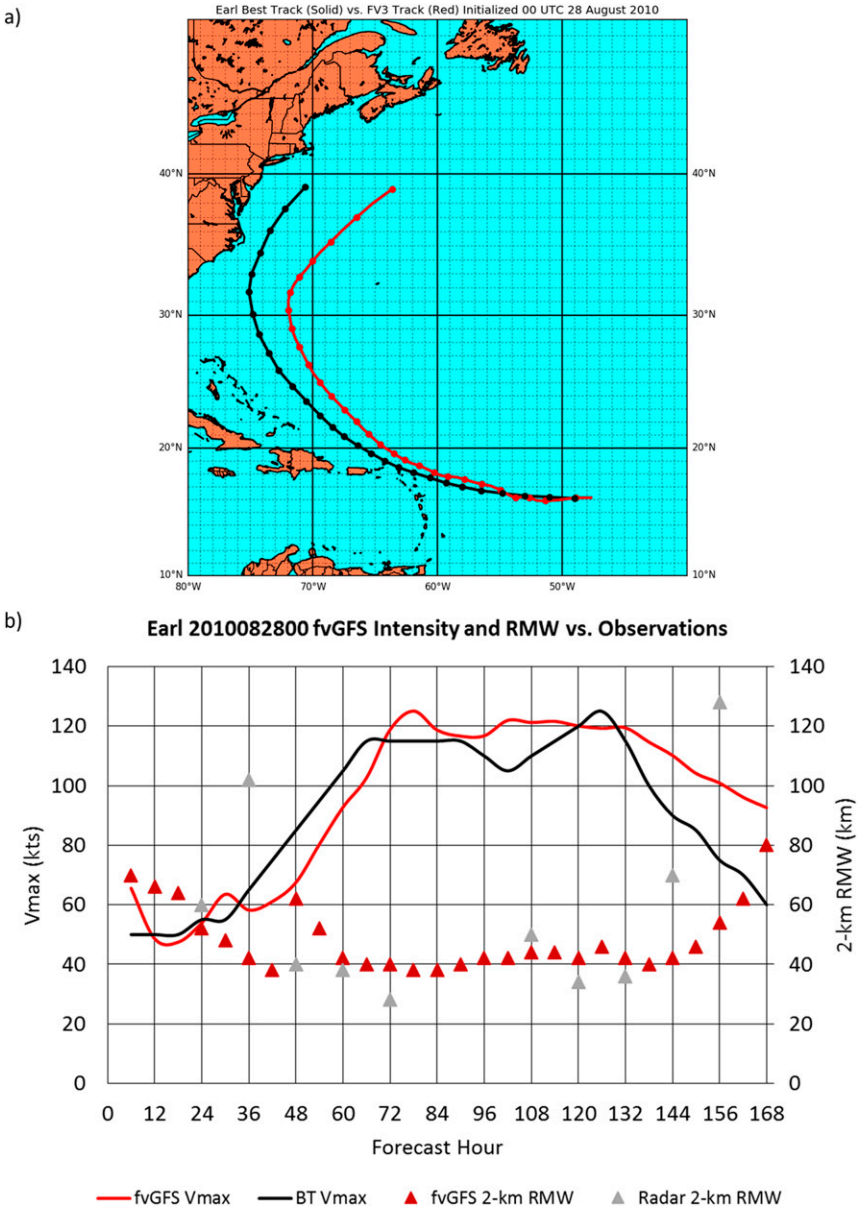

FIG. 6. (a) The track of Hurricane Earl starting at 0000 UTC 28 Aug 2010 and ending at 0000 UTC 4 Sep 2010 from the NHC best-track dataset (black, every 6 h) and the fvGFS forecast (red, every 6 h). (b) Intensity of Hurricane Earl starting at 0000 UTC 28 Aug 2010 and ending at 0000 UTC 4 Sep 2010 from the NHC best-track dataset (black) and the fvGFS forecast (red). The 2-km RMW forecasts from fvGFS are shown by the dark red triangles and observed 2-km RMW based on radar data are shown by the gray triangles.

decayed more slowly than the observed TC. The model TC structure is next compared with radar structure at a couple of select points in the life cycle of the TC.

Figure 7 shows the 2-km and azimuthal-mean tangential wind from the radar data from 2057 UTC 29 August to 0038 UTC 30 August. The 48-h forecast tangential wind from fvGFS is also shown. The general structure from the model looks very similar to the observations horizontally, with a peak in the inner eyewall and a second peak in a band farther radially outward. The vertical structure was not as well predicted, however. Looking at some of the individual stats at this time, the model RMW was 62 km versus the observed 40 km, and the calculated model vortex depth was 9.25 km compared to the observed 7 km. Despite the calculated larger vortex depth (probably due to the model winds around 30 m s^{-1} extending well into the

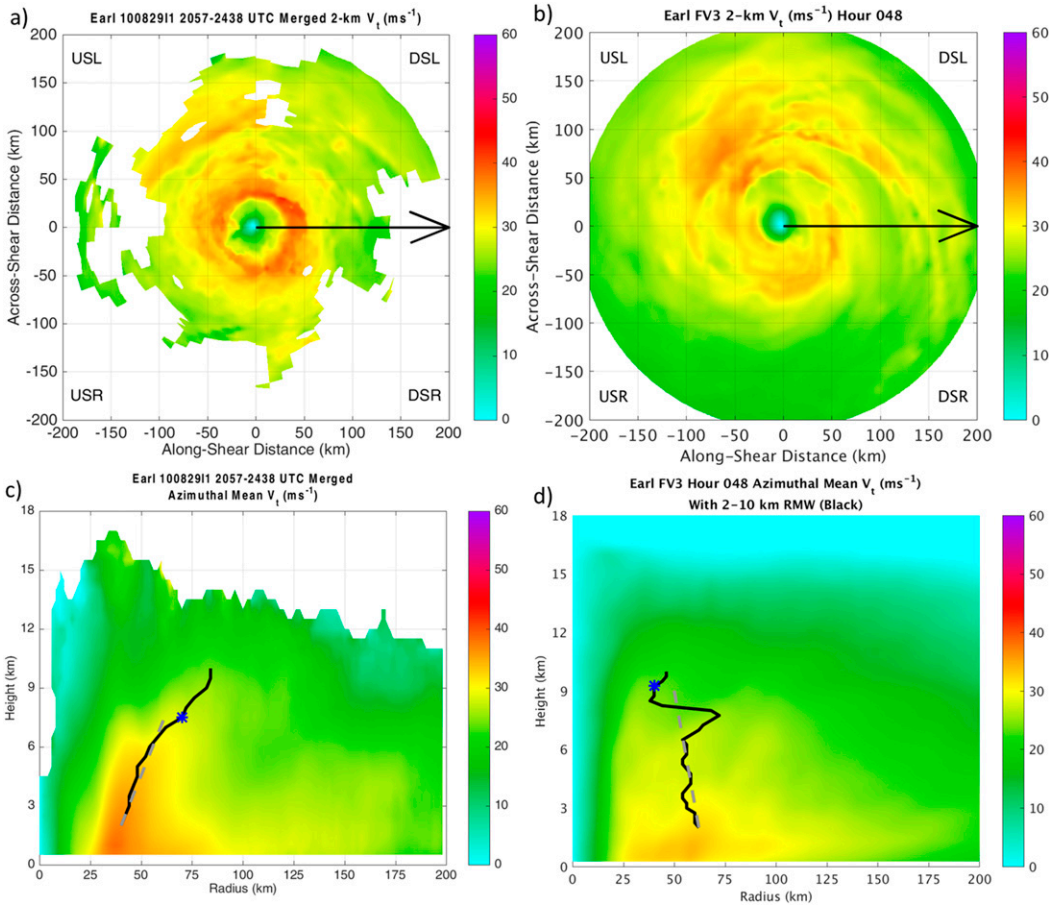

FIG. 7. (a) The 2-km tangential wind (m s^{-1}) from the P-3 flight into Hurricane Earl from 2057 UTC 29 Aug to 0038 UTC 30 Aug 2010. The black arrow represents the shear vector, and the reference frame is rotated to a shear-relative coordinate system. (b) As in (a), but for the 48-h fvGFS forecast, valid at 0000 UTC 30 Aug 2010. (c) Azimuthal-mean tangential wind (m s^{-1}) from the P-3 flight into Hurricane Earl from 2057 UTC 29 Aug to 0038 UTC 30 Aug 2010. The black line is the azimuthal-mean RMW, the blue star represents the calculated vortex depth, and the dashed gray line is the calculated RMW slope. (d) As in (c), but for the 48-h fvGFS forecast, valid at 0000 UTC 30 Aug 2010.

troposphere), the strongest winds in the model are confined to a shallower layer than in the radar observations, and the model vortex does not appear to be as well defined as the observed. From Fig. 7, it also appears that the model RMW slope is more upright than in the observations as a result of the RMW "jumping" around vertically within a relatively broad wind peak due to the banding structure.

Figure 8 shows the 2-km and azimuthal-mean tangential wind 24 h later, with a flight from 2112 to 2334 UTC August 30 and the model data valid at 0000 UTC 30 August (a 72-h forecast). The observed TC had deepened by 30 kt during this period and then had begun a period of steady intensity, likely due to the formation of the secondary eyewall (evident in Fig. 8a around a radius of 100 km). The model TC had actually deepened by about 50 kt during the same period and was slightly too strong. This is evident in the magnitude of the tangential wind at $z = 2$ km. The RMW was still slightly too large (40 vs 28 km in the observations), but the vortex depth (12 km in both the model and observations) and RMW slope (0.37 in the model and −0.21 in the observations) indicated a deep, upright vortex both in the real storm and in the model forecast. The differences in the secondary wind maximum are reflected in the vortex decay parameter. For the model, $\alpha = 0.54$, while for the radar data, $\alpha = 0.23$, indicating a much broader wind peak in the observed TC. The model did show a slight secondary wind peak around hours 78–84 (not shown because there were no radar data for direct comparison), with α decreasing to 0.49 and the intensity leveling off (Fig. 6). However, the wind peak was still too sharp, and there was not an appreciable increase in

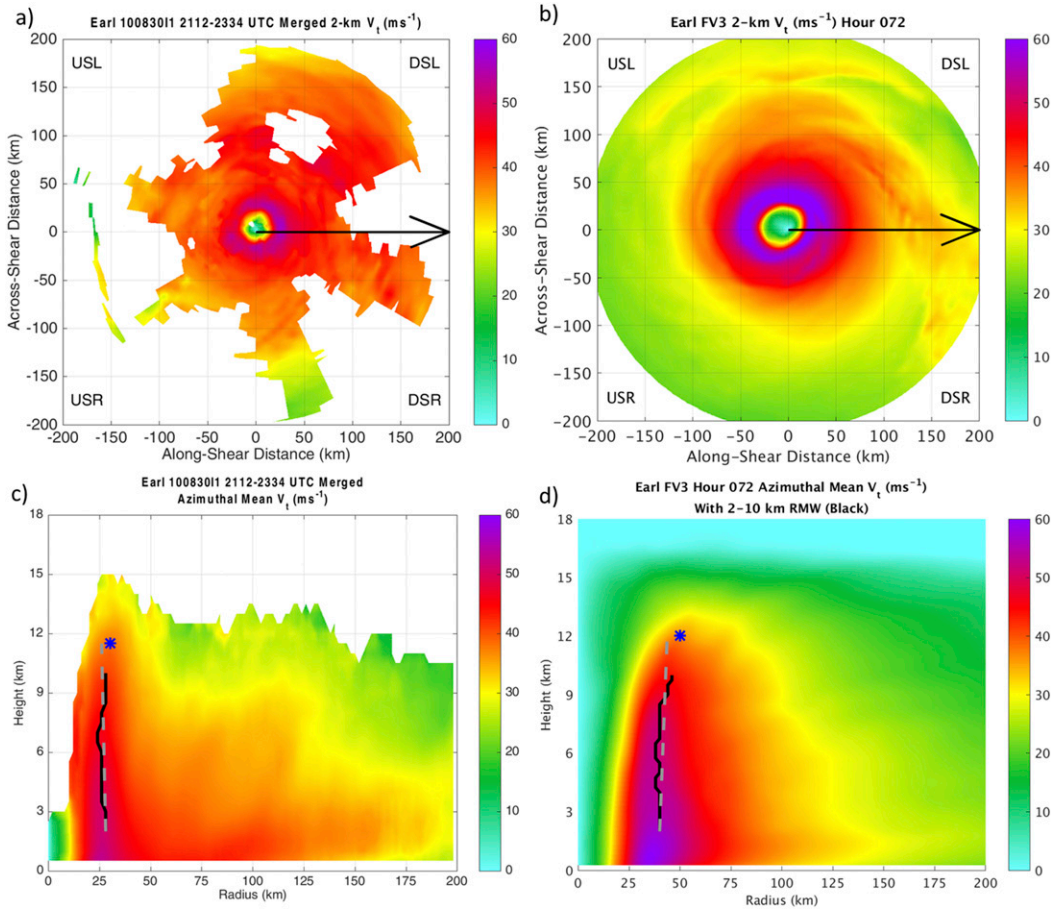

FIG. 8. As in Fig. 7, but the flight was from 2112 to 2334 UTC 30 Aug, and the model data are a 72-h forecast, valid at 0000 UTC 31 Aug.

the RMW, which is typically a hallmark of observed eyewall replacement cycles (e.g., Willoughby et al. 1982). Thus, the model's overforecast of intensity during this period was likely tied to the incorrect depiction of the secondary eyewall.

b. *Hurricane Danny (2015) initialized 1200 UTC 19 August 2015*

Another case study focused on a forecast of Hurricane Danny (2015). Danny was a low-latitude "midget tropical cyclone" (e.g., Harr et al. 1996) that experienced both rapid intensification and rapid weakening. The TC deepened by 65 kt from 1200 UTC 19 August to 1200 UTC 21 August, developing from a tropical storm with 45-kt wind into a 110-kt major hurricane. After this, the TC quickly weakened as a result of increasing vertical shear and the intrusion of dry air into the core (Stewart 2016), and by 1200 UTC 23 August it was once again a 45-kt tropical storm. The fvGFS forecast of Danny was initialized at 1200 UTC 19 August and run for 5 days (120 h).

The track and intensity of Danny from the best-track data and the fvGFS forecast are shown in Fig. 9. The model track was close to observations initially, but then was slightly south of the actual storm. The trajectory of the modeled storm was approximately west-northwest throughout most of the forecast, while the observed storm moved northwest around 50°W before turning to a more westerly heading. The model did not quite capture the peak in intensity (perhaps because of the resolution), but did come close to capturing the intensification rate during RI, with an intensification of 34 kt between hours 30 and 54. The model also did not capture the sharp dropoff in intensity, with only slow weakening from hours 60 to 96 before a weakening trend began in the final 24 h of the forecast. The first flight was at the time of observed peak intensity, so unfortunately the RMW evolution during the intensification could not be observed. However, the RMW expanded during the weakening period. The model RMW shrank by ~30 km during the intensification period and did expand slightly

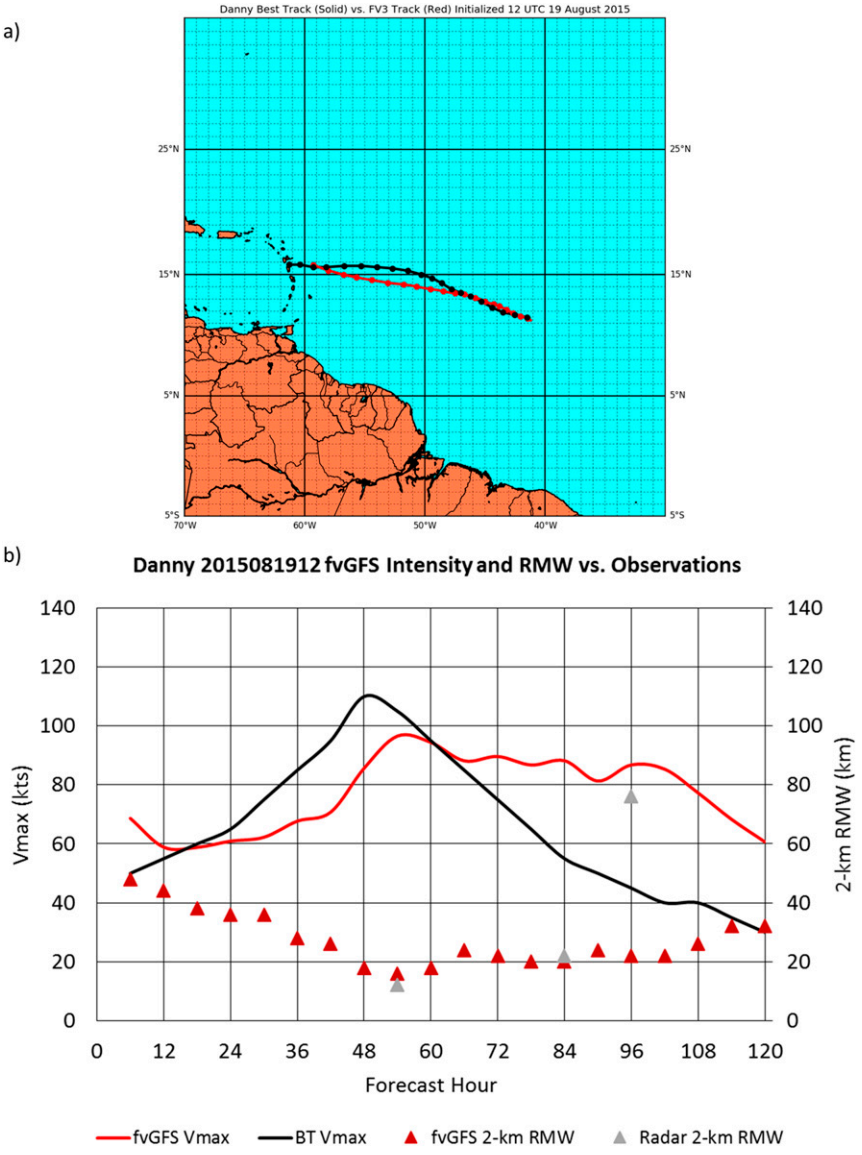

FIG. 9. As in Fig. 6, but for the (a) track and (b) intensity of Hurricane Danny starting at 1200 UTC 19 Aug 2015 and ending at 1200 UTC 24 Aug 2015.

(by about 15 km) during the weakening, but not as quickly as in the observed TC.

The track and intensity errors in this forecast were likely connected. The fact that the model TC was weaker than the observed storm during RI likely led to the modeled TC moving west-northwest less quickly. This can be seen in Fig. 10a, which shows the GFS analysis of layer-mean steering winds in six layers between 850–700 and 850–200 hPa at 1200 UTC 21 August. The layer means are calculated from the environmental wind in a $10° \times 10°$ box around the TC, with the $2° \times 2°$ box closest to the center removed to eliminate the TC itself. At this time, the observed TC SLP was 960 hPa, and the SLP of the forecast TC was 977 hPa. The deeper-layer flow felt by the stronger observed TC (Velden 1993) was more to the west-northwest than the shallower flow, which was more zonal. Later, however, as the modeled TC did not decay quickly enough, it moved more north of west than the observed TC, which weakened quickly and moved nearly due west. In addition, the differences in the rate of weakening are connected to the track difference leading up to peak intensity, as is shown in Figs. 10b,c. Danny was moving toward a sharp shear axis associated with the tropical upper-tropospheric trough (TUTT; Sadler 1976; Fitzpatrick et al. 1995), which had a significant weakening effect on the TC. The observed

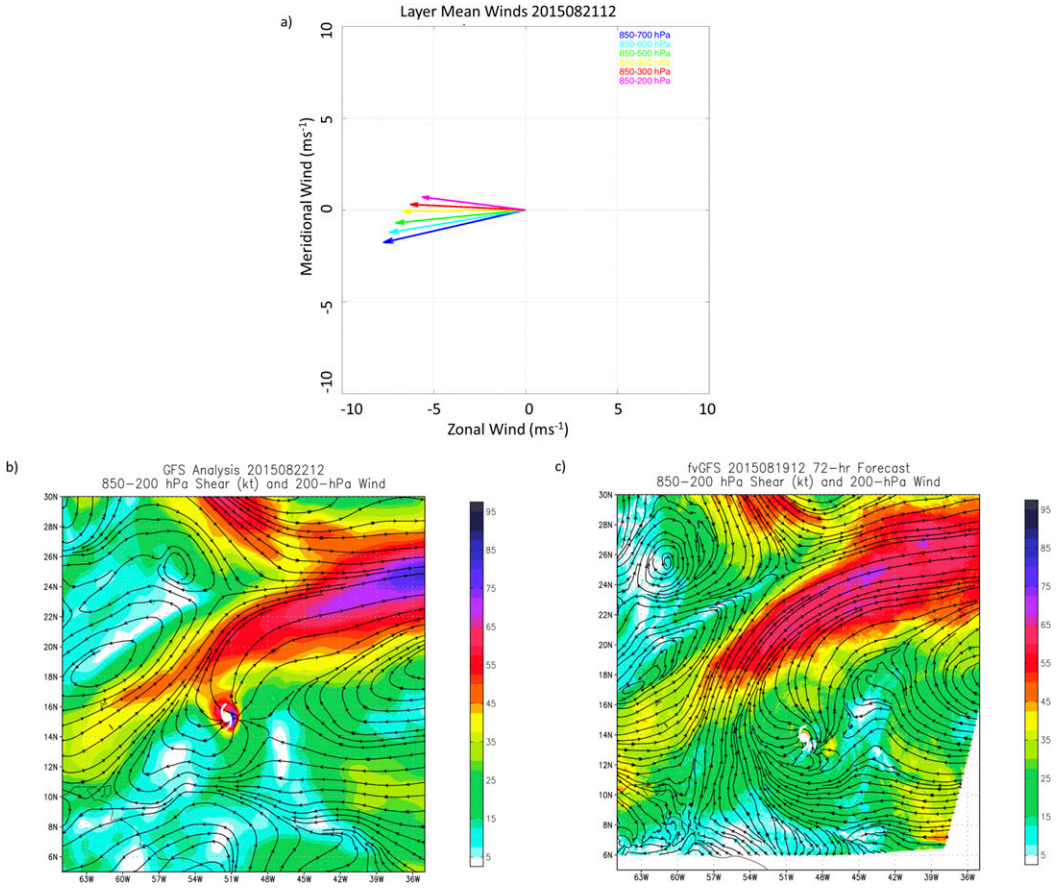

FIG. 10. (a) The GFS analysis layer-mean winds from six different layers (see legend) at 1200 UTC 21 Aug 2015. (b) The GFS analysis 850–200-hPa shear (kt; shaded) and 200-hPa wind (streamlines) at 1200 UTC 22 Aug 2015. The position of Hurricane Danny is shown with the white hurricane symbol. (c) As in (b), but for the fvGFS 72-h forecast valid at 1200 UTC 22 Aug 2015.

storm, which moved northwest more quickly, encountered this shear axis sooner than the modeled TC, which stayed slightly farther southeast (around 14°N and 49°W compared to the observed TC around 15°N and 51°W), allowing it to avoid the shear axis until later in the forecast. The modeled TC also appeared to have a larger anticyclone aloft (Fig. 10c), insulating it from the shear longer.

Figures 11a and 11c show the 2-km and azimuthal-mean tangential wind from the radar data from 1620 to 1705 UTC 21 August 2015, along with the tangential wind from the 54-h fvGFS forecast (Figs. 11b,d). While the inner core is not quite as tiny in the model as in the observations (perhaps because of the resolution), the model does capture the very small nature of the storm. The RMW was 12 km in the radar data and 16 km in the model. In addition, the radar α (0.62) was close to the forecast α (0.65), indicating a sharp wind peak in both the model and observations. The model RMW slope (1.97) was greater than in the observations (0.62), but the model vortex depth was greater (7 vs 7.75 km). It appears that the model forecasted a slightly stronger deep-layer vortex than was observed, perhaps because of the excessive PBL depth (e.g., Gopalakrishnan et al. 2013). This depth difference is perhaps counterintuitive given that the modeled storm appeared to feel less impact from deep-layer shear, indicating that the proximity to the shear axis was a key difference, with the observed TC closer and affected earlier.

Figure 12 shows the observed reflectivity at 5 km from the flight as well as the model-simulated 5-km reflectivity at hour 54, both in a shear-relative coordinate system. While the actual reflectivity values may not be perfect, especially in the observations because of the merging technique, the structure can be reasonably compared. Both TCs have a very small eye and inner-core region, with small bands in the downshear-left region. The radar eye appears smaller, although it may have been obscured somewhat by the merging of flight legs. One big difference is that the modeled TC has fairly

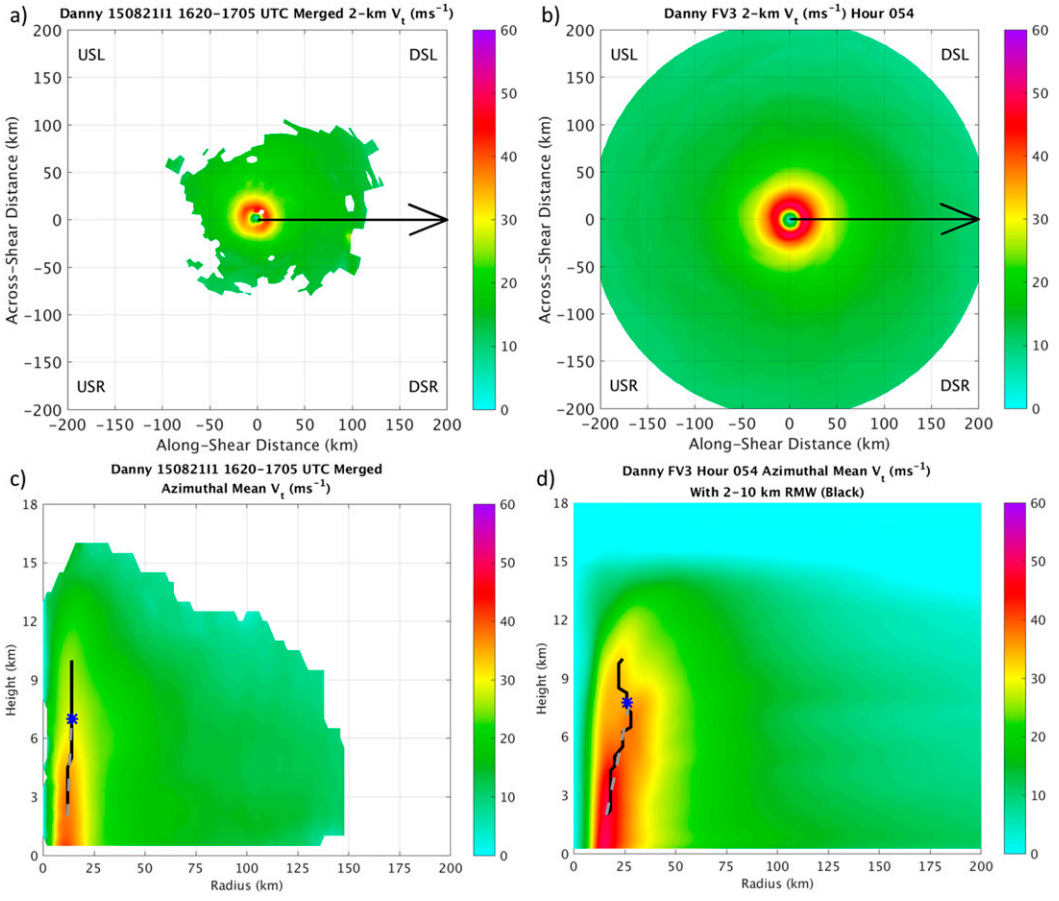

FIG. 11. (a) The 2-km tangential wind (m s^{-1}) from the P-3 flight into Hurricane Danny from 1620 UTC 21 Aug to 1705 UTC 21 Aug 2015. The black arrow represents the shear vector, and the reference frame is rotated to a shear-relative coordinate system. (b) As in (a), but for the 54-h fvGFS forecast, valid at 1800 UTC 21 Aug 2015. (c) Azimuthal-mean tangential wind (m s^{-1}) from the P-3 flight into Hurricane Danny from 1620 UTC 21 Aug to 1705 UTC 21 Aug 2015. The black line is the azimuthal-mean RMW, the blue star represents the calculated vortex depth, and the dashed gray line is the calculated RMW slope. (d) As in (c), but for the 48-h fvGFS forecast, valid at 1800 UTC 21 Aug 2015.

symmetric precipitation, while the radar shows more of a downshear-left maximum. This is further evidence that the model TC did not react to the shear as quickly as the real TC. Despite this difference, the small eye shows that the model is capable of resolving inner-core precipitation and wind structure even in very small systems.

c. *Hurricane Matthew initialized 0000 UTC 5 October 2016*

The final case study examines the second forecast of Hurricane Matthew. Hurricane Matthew moved through the Caribbean Sea and then traveled across western Haiti, eastern Cuba, and the Bahamas before coming very close to the eastern coast of Florida and much of the southeast U.S. coastline of Georgia and the Carolinas. After moving off of the Cuba coast, Matthew remained approximately steady in intensity for ~24 h as the core recovered from land interaction. It then intensified by 20 kt in 18 h over the warm waters in the Bahamas, before beginning to weaken due to an ERC (Stewart 2017). The TC then began to weaken more quickly as it was impacted by vertical wind shear and an intrusion of continental dry air.

The fvGFS forecast of Matthew was initialized at 0000 UTC on 5 October, as the TC was beginning to move away from Cuba. Figure 13 shows the track, intensity, and RMW of the model forecast versus the observations for the 5-day forecast. The track is generally very consistent with the observations, except for a northward bias during the recurve/extratropical transition phase as the TC decayed. The intensity evolution is also generally consistent with the observations, with a peak of just less

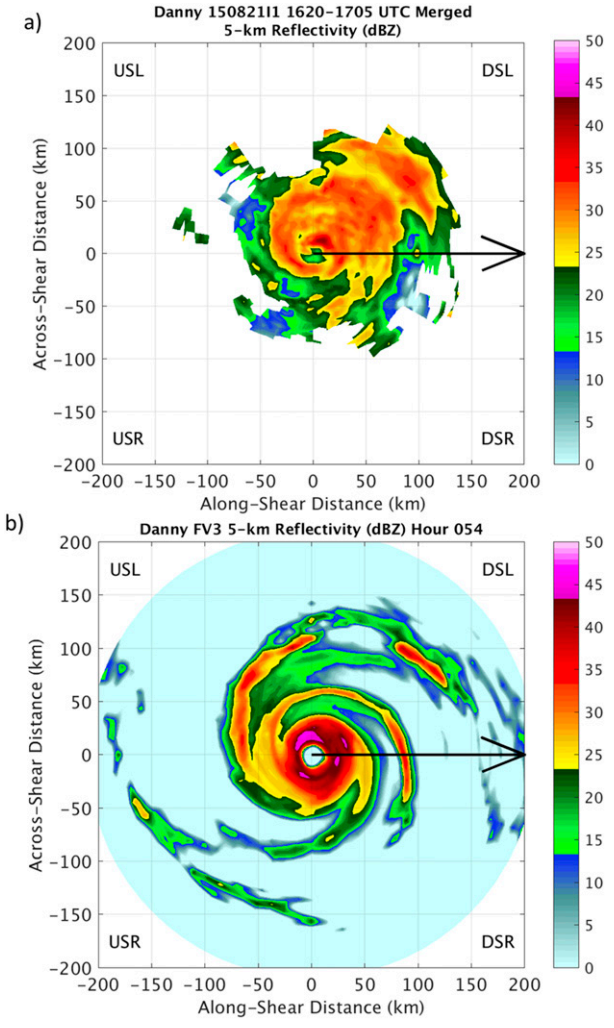

FIG. 12. (a) The 5-km reflectivity (dBZ) from the P-3 flight into Hurricane Danny from 1620 UTC 21 Aug to 1705 UTC 21 Aug 2015. The black arrow represents the shear vector, and the reference frame is rotated to a shear-relative coordinate system. (b) As in (a), but for the 54-h fvGFS forecast of 5-km simulated reflectivity (dBZ) valid at 1800 UTC 21 Aug 2015.

than 120 kt at hour 42, followed by a weakening period, which appears to be due to the model correctly forecasting the ERC (which will be discussed later). However, the modeled storm has an unrealistic reintensification period after the ERC is completed, from hours 54 to 66. Beyond 66 h, both the model and observations show a steady to rapid decay in intensity for the rest of the period. The RMW forecast is also generally consistent with the observations. After a ~12-h spinup period (since the GFS initial conditions could not resolve the inner core of a major hurricane), the model RMW decreases slowly as the TC intensifies, then increases as a result of the secondary eyewall formation (as well as poleward movement). After completing the ERC, the model RMW contracts slightly during the (nonobserved) reintensification period before increasing sharply during the decay phase and extratropical transition. Although the observational radar dataset did not cover this period, the RMW from the best-track dataset also showed a sharp increase during this period, from 46 km at 1800 UTC 8 October 8 to 111 km at 1800 UTC 9 October. The model TC did appear to complete the ERC somewhat faster than the observed TC, which is why the model TC intensified before the shear increased.

Figure 14 shows the 2-km and azimuthal-mean tangential winds from 0954 to 1245 UTC 5 October, as well as the 12-h fvGFS forecasts of the tangential wind. The model RMW was 38 km, only slightly larger than the 32 km observed in the radar data. The wind peak in the model is slightly sharper than in the observations, with $\alpha = 0.49$ in the radar data at $z = 2$ km and $\alpha = 0.57$ in the 12-h fvGFS forecast. Looking at the vertical structure parameters, the vortex depth is 6.5 km in the radar data and 7 km in the model forecast. This depth is unusually shallow for a TC that was still a major hurricane at the time, which is perhaps related to the convection weakening as a result of land interaction. The RMW slope in the model was 1.75, versus 2.31 in the radar data. Despite slight differences, which are to be expected with the model spinning up from the GFS initial conditions, the short-term structure appears to be well forecast in this case.

Figure 15 shows the azimuthal-mean and 2-km tangential wind 36 h later, at hour 48 of the forecast. At this time, the developing secondary wind maximum can be seen in the radar data at a radius of 50–75 km, although the inner eyewall still appears to be dominant. In the model data, two features stand out. First, the tangential wind is stronger throughout the depth of the vortex in the model. Second, although it is not as apparent as in the radar data and is more like a primary band structure, the model does show a secondary wind maximum developing around $r = 75$ km. The secondary maximum appears in several fields, including reflectivity (Fig. 16a), vertical velocity (Fig. 16b), 850-hPa wind speed (Fig. 16c), and 850-hPa vorticity (Fig. 16d). In the model, some small concentric rings and moat regions are seen just outside the inner eye, but the most prominent secondary maximum is connected to a band spiraling into the storm. The relative scales of the two eyewalls are generally similar to those seen in a WRF simulation of secondary eyewalls during Hurricanes Katrina and Rita (2005) by Abarca and Corbosiero (2011). The strength of the secondary maximum is reflected in the vortex decay parameter, with $\alpha = 0.17$ in the radar data and $\alpha = 0.40$ in the model data. The RMW is also smaller in the radar (20 km) than in the forecast (32 km). Although

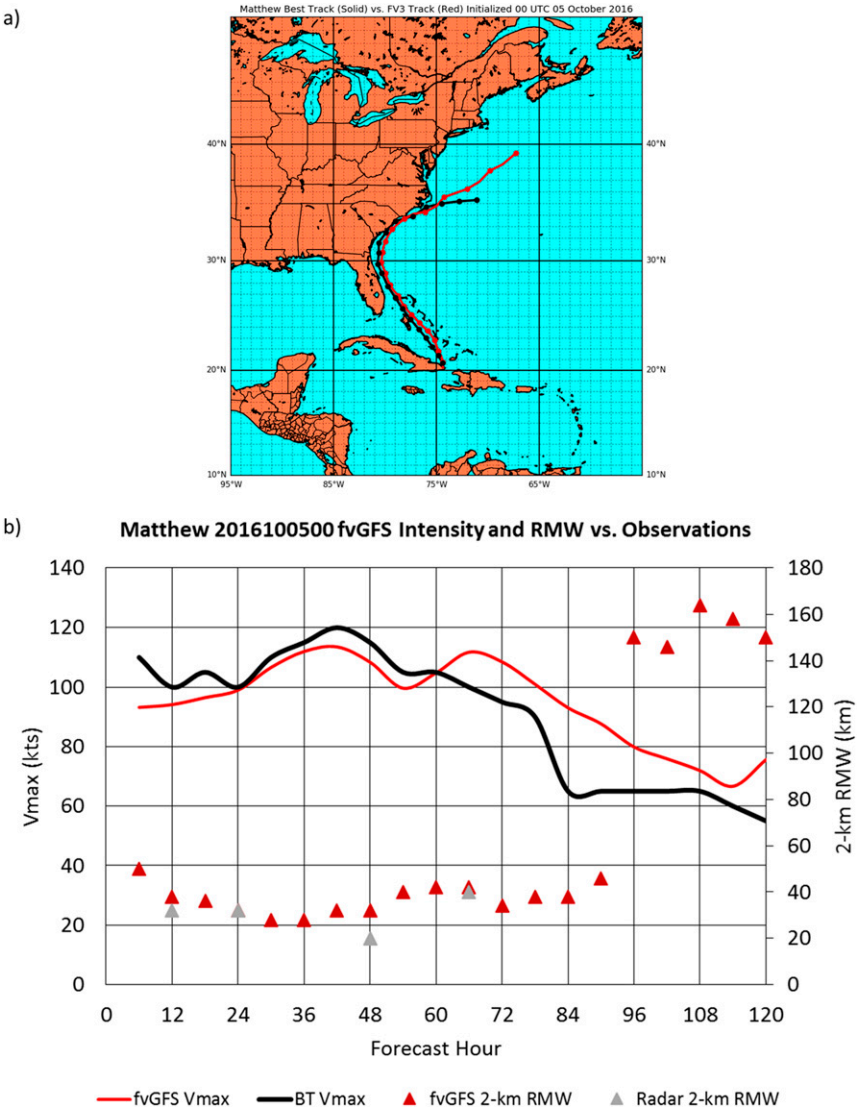

FIG. 13. As in Fig. 6, but for the (a) track and (b) intensity of Hurricane Matthew starting at 0000 UTC 5 Oct 2016 and ending at 0000 UTC 10 Oct 2016.

the inner eyewall is relatively upright in both the model and observations (slope of 0.77 in the model data and 0.49 in the radar data), the vortex is too deep in the forecast (12 km) compared to the radar data (9 km). The model intensity does decrease slightly after this time, along with an increase in RMW, which are both consistent with an eyewall replacement cycle. However, the model appears to complete the ERC too quickly, leading to the mentioned period of unrealistic reintensification (Fig. 13).

5. Conclusions

The composite results and case studies shown in this study demonstrate the effectiveness of airborne Doppler radar data as a tool for model evaluation, thanks to its high resolution and thorough coverage of the TC inner core. In addition, the promising track, intensity, and structure forecasts from fvGFS show that, although the model is still in the early stages of development, it has the ability to successfully predict TCs at high resolution. In particular, the ability to generate a secondary eyewall and eyewall replacement cycle in the Matthew forecast highlighted the ability of fvGFS to predict some of the structural changes that are critical for forecasting TC intensity and impact. This gives confidence that skillful model track and intensity forecasts are right for physically realistic reasons, although the structure is still imperfect and more work is needed. These findings motivate further improvements in structure

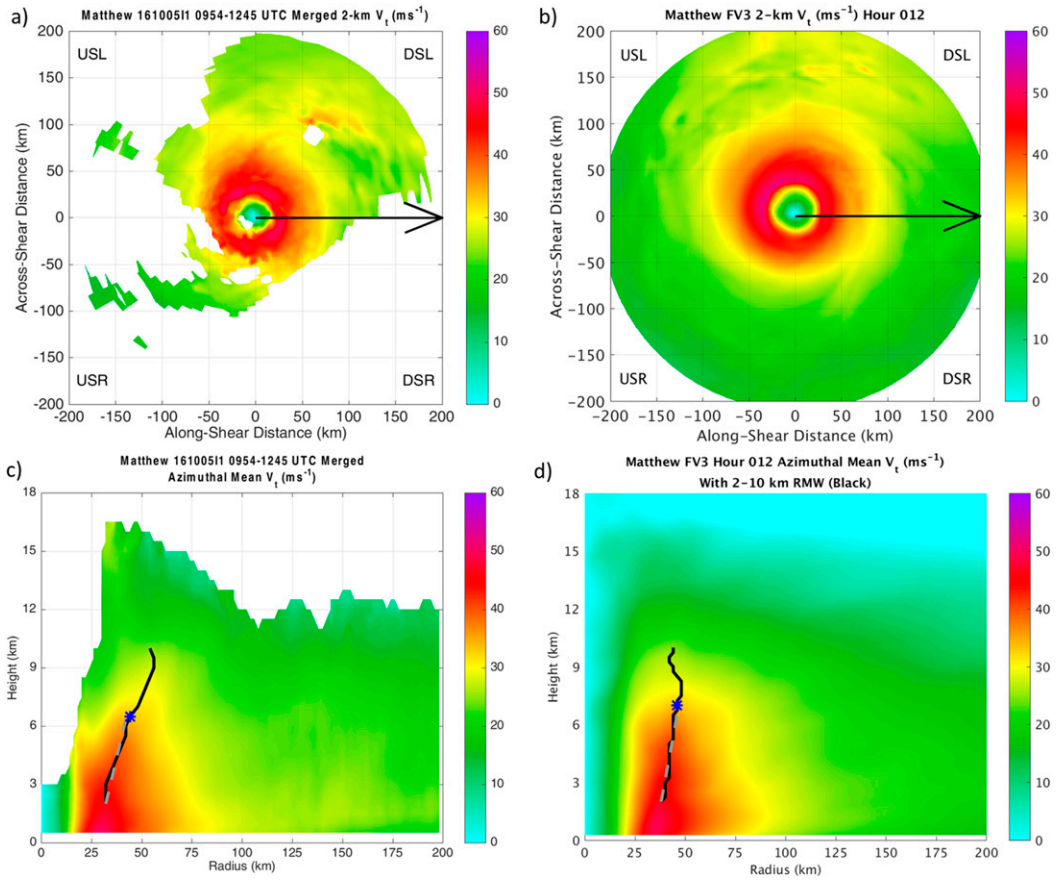

FIG. 14. (a) The 2-km tangential wind (m s^{-1}) from the P-3 flight into Hurricane Matthew from 0954 UTC 5 Oct to 1245 UTC 5 Oct 2016. The black arrow represents the shear vector, and the reference frame is rotated to a shear-relative coordinate system. (b) As in (a), but for the 12-h fvGFS forecast, valid at 1200 UTC 6 Oct 2016. (c) Azimuthal-mean tangential wind (m s^{-1}) from the P-3 flight into Hurricane Matthew from 0954 UTC 5 Oct to 1245 UTC 5 Oct 2016. The black line is the azimuthal-mean RMW, the blue star represents the calculated vortex depth, and the dashed gray line is the calculated RMW slope. (d) As in (c), but for the 12-h fvGFS forecast, valid at 1200 UTC 6 Oct 2016.

forecasts and, by extension, the prediction of basic track and intensity metrics. Further upgrades to fvGFS and other models can be enhanced by changes that lead to the enhanced ability to predict this kind of finescale structure. The track and intensity errors are generally close to the mean NHC forecast errors during the period observed, although there is a 6–12-h "spinup" process due to the low-resolution initial conditions that can lead to errors depending on how quickly the model spins up the TC vortex. The bias toward being too weak at short lead times flips to a bias toward being too intense at longer lead times, perhaps because of the lack of ocean coupling in the model.

The individual structure metrics highlight some consistent patterns within the model TC forecasts. The RMWs tend to be too small for very large observed RMWs, which is perhaps associated with the bias toward TCs that are too strong. In addition, the horizontal wind peak is typically too sharp, which is also consistent with the high bias in intensity (Mallen et al. 2005). In terms of vertical structure, the vortex tends to be somewhat too deep when the observed TC is shallow, although deep storms are typically well predicted. The eyewall slope relationship is not as clear, but the model seems to have a preferred range from around 0.5 to 2, which does not capture many of the observed very upright or very slanted eyewalls. Based on a comparison of different error metrics, the RMW was generally the best predicted of the four main structure variables, although the results are not consistent between the metrics. In addition, some of the observed relationships between different structural features (both in the literature and this dataset), such as the correlation between the eyewall slope and RMW, were not well reproduced by the model.

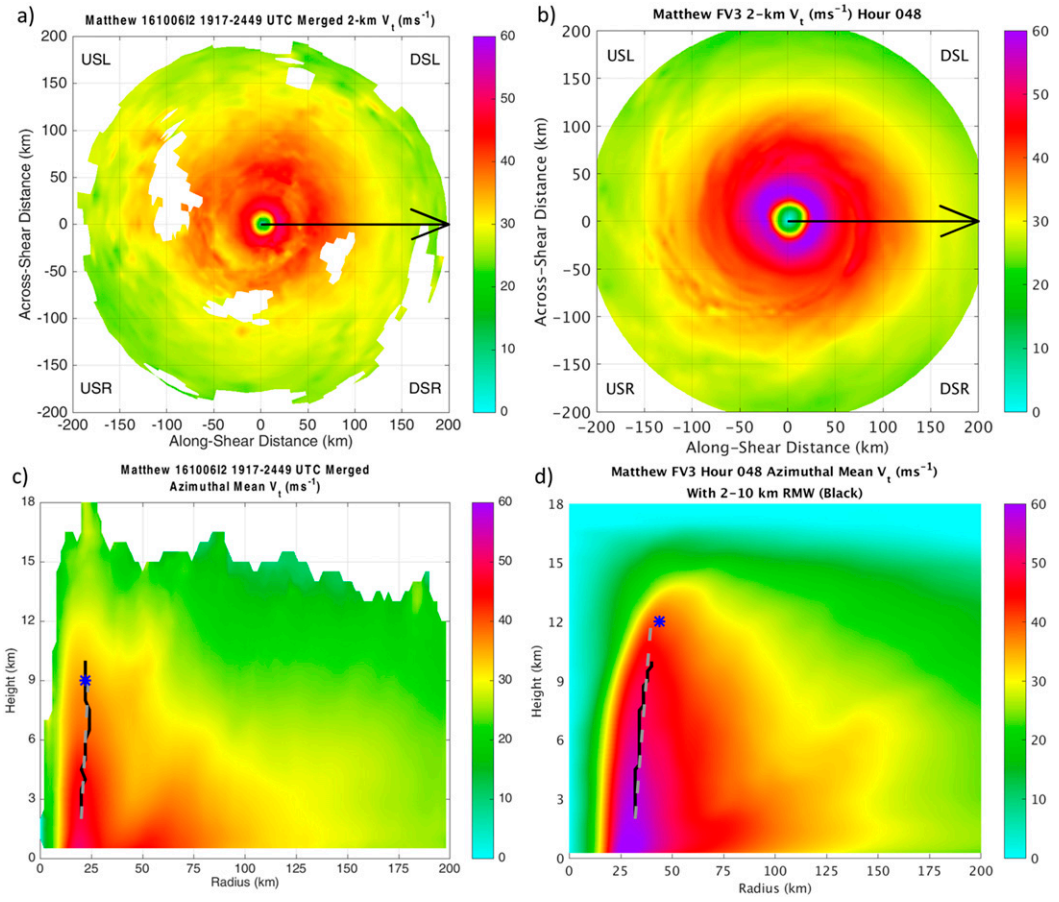

FIG. 15. As in Fig. 14, but here the flight was from 1917 UTC 6 Oct to 0049 UTC 7 Oct, and the model data are a 48-h forecast, valid at 0000 UTC 7 Oct.

The analysis of the composite structures in the model, and comparisons with homogeneous composites from the airborne radar data, provide insight into the mean structure predicted by the model over a range of lead times. The tangential wind composite is generally consistent with observations, although the vortex is too deep and has a sharper wind peak in the model. The radial flow composites reveal that the boundary layer inflow tends to be too deep in the model but that the outflow aloft is consistent with the observational composite. The former could likely be substantially improved with an upgraded PBL scheme. The shear-relative composites of radial flow highlight the fact that the model produces mostly realistic shear-relative wavenumber-1 asymmetries in the TCs (with some slight differences in the downshear-left region).

Individual case studies build on the composites and statistical results through examining the model TC structure by providing examples of the types of structures the model is capable of predicting and comparing these directly with the radar data. For Hurricane Earl, the model was able to approximately reproduce the observed rapid intensification. In addition, the deep, upright eyewall near peak intensity in the model forecast was validated by the airborne Doppler data. However, the model did not capture the steady/weakening period due to an eyewall replacement cycle and also remained somewhat too strong during the recurvature part of the track. For Danny, the model was able to show the period of rapid intensification with relatively good timing, although the peak intensity was slightly less than observed. The model also was able to come close to capturing the tiny inner core of Danny, with a model RMW of 16 km (not far from the observed 12 km). Unfortunately, however, Danny was another case where the vortex stayed too strong and deep later in the forecast period, which in this case appeared to be due to a track error leading to a later interaction with hostile wind shear. Finally, in the Matthew case, the model accurately predicted the horizontal and vertical wind structure and intensity after the storm moved away from Cuba, although the vortex was somewhat too deep and

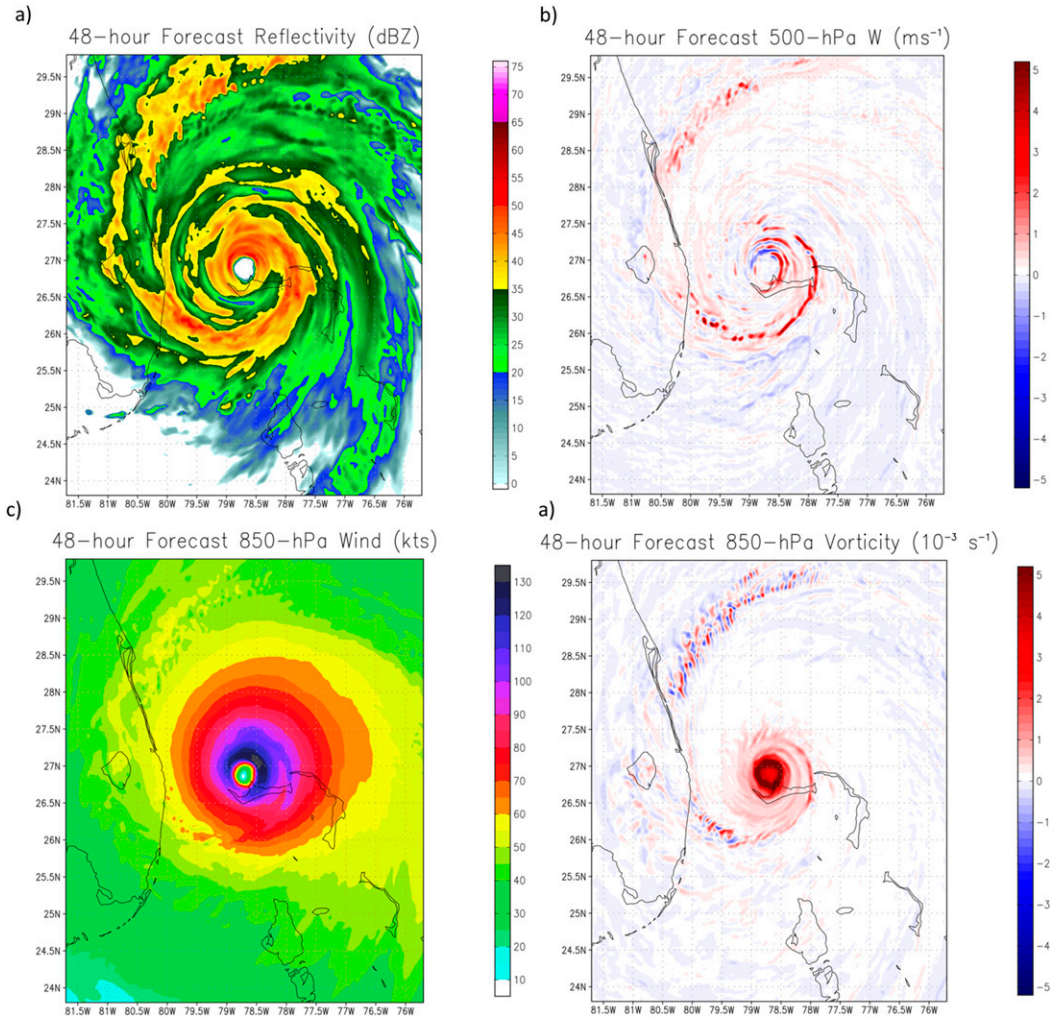

FIG. 16. The fvGFS 48-h forecast (a) composite reflectivity (dB*Z*), (b) 500-hPa vertical velocity (m s^{-1}), (c) 850-hPa wind speed (m s^{-1}), and (d) 850-hPa vorticity ($\times 10^{-3}$ s^{-1}) of Hurricane Matthew, valid at 0000 UTC 7 Oct 2016.

intense. The model was also able to generate a secondary eyewall in this forecast at approximately the correct time, with about ~48-h lead. However, differences in the relative size and scale of the two eyewalls (as compared with the observed airborne radar data) could explain why the model storm reintensified after the ERC while the observed storm did not. Taken as a whole, these case studies indicate that radar data serve as a valuable verification tool for three-dimensional forecast TC structure. In addition, the radar results indicate that fvGFS shows promise in its ability to forecast TC structure and intensity change at high resolution. As models continue to improve in resolution and physics, further evaluation of the finescale core structure will be a valuable tool for helping to improve predictions.

Future work will continue to explore TC structure in fvGFS and examine new ways to use the airborne Doppler and other datasets to analyze forecasts of TC structure. We currently have a small number of forecasts; therefore, a rigorous analysis of the time-dependent biases is not feasible. For example, the first Hurricane Hermine forecast had an RMW of over 150 km at 0000 UTC 1 September 2016 (a 48-h forecast), but the forecast initialized a day later had an RMW of 38 km in its 24-h forecast. For the Matthew case, however, the 7-day forecast from the early run had a relatively similar RMW to a 1-day forecast from a later run at the same valid time (42 vs 32 km). It is not clear whether the inconsistent valid times here impacted the results, but we plan to perform a larger number of retrospective forecasts to make our analysis more rigorous and to have a larger diversity of storms. It would also be useful to evaluate weaker TCs, as all of the storms in this study achieved at least hurricane intensity. Although radar coverage is often

poor in disorganized storms, an evaluation of track and intensity prediction over a wider range of intensities would prove insightful, especially if it included an evaluation of model forecasts of vortex tilt in sheared environments. The current study focused mostly on the examination of horizontal wind structure from radar data. Other remotely sensed data (such as satellite winds and/or rainfall data) would also be useful for verification of the model's dynamic and thermodynamic structure forecasts.

Finally, it would likely prove insightful to evaluate different combinations of model physics and use this kind of structural evaluation to aid in model development. For example, a different boundary layer scheme in development for use in fvGFS could lead to improvements in the 10-m wind and low-level inflow structure, and this could be tested using a larger composite such as the one shown here. Although beyond the scope of this study, the vertical velocity and precipitation structures are likely sensitive to the cumulus parameterization and microphysics schemes, and as fvGFS is able to use different schemes, this radar dataset will provide an excellent source for comparison of this structure. In addition, as an ocean coupling capability is developed, it will be useful to make sure this change positively impacts the vortex depth and wind structure, not just the maximum 10-m wind.

Acknowledgments. Paul Reasor provided invaluable support through instruction on how to create the three-dimensional Doppler analyses. We thank the crews of the NOAA-P3 aircraft for their tireless dedication to collecting these data. Morris Bender, Gus Alaka, and three anonymous reviewers provided helpful comments and discussions that led to improvements on an earlier version of the manuscript. ATH was supported by NOAA Grant NA14OAR4830101.

REFERENCES

Abarca, S. F., and K. L. Corbosiero, 2011: Secondary eyewall formation in WRF simulation of Hurricanes Rita and Katrina (2005). *Geophys. Res. Lett.*, **38**, L07802, https://doi.org/10.1029/2011GL047015.

Aberson, S. D., 1998: Five-day tropical cyclone track forecasts in the North Atlantic basin. *Wea. Forecasting*, **13**, 1005–1015, https://doi.org/10.1175/1520-0434(1998)013<1005:FDTCTF>2.0.CO;2.

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.*, **31**, 674–701, https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

Bender, M. A., 2017: Evaluation of tropical cyclone forecasts with the GFDL FV3 dynamical core. *71st Interdepartmental Hurricane Conf.*, Miami, FL, Federal Coordinator for Meteorological Services and Supporting Research, http://www.ofcm.gov/meetings/TCORF/ihc17/Session_08/8-3-Bender_web.pdf.

Black, P. G., and Coauthors, 2007: Air–sea exchange in hurricanes: Synthesis of observations from the Coupled Boundary Layer Air–Sea Transfer Experiment. *Bull. Amer. Meteor. Soc.*, **88**, 357–374, https://doi.org/10.1175/BAMS-88-3-357.

Cangialosi, J. P., 2011: Tropical cyclone report Hurricane Earl (AL072010) 25 August–4 September 2010. National Hurricane Center Rep., 29 pp., http://www.nhc.noaa.gov/data/tcr/AL072010_Earl.pdf.

Chen, H., and S. G. Gopalakrishnan, 2015: A study on the asymmetric rapid intensification of Hurricane Earl (2010) using the HWRF system. *J. Atmos. Sci.*, **72**, 531–550, https://doi.org/10.1175/JAS-D-14-0097.1.

Chen, J.-H., and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate*, **26**, 380–398, https://doi.org/10.1175/JCLI-D-12-00061.1.

DeHart, J. C., R. A. Houze Jr., and R. F. Rogers, 2014: Quadrant distribution of tropical cyclone inner-core kinematics in relation to environmental shear. *J. Atmos. Sci.*, **71**, 2713–2732, https://doi.org/10.1175/JAS-D-13-0298.1.

DeMaria, M., and J. Kaplan, 1994: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220, https://doi.org/10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2.

Fierro, A. O., R. F. Rogers, F. D. Marks, and D. S. Nolan, 2009: The impact of horizontal grid spacing on the microphysical and kinematic structures of strong tropical cyclones simulated with the WRF-ARW Model. *Mon. Wea. Rev.*, **137**, 3717–3743, https://doi.org/10.1175/2009MWR2946.1.

Fitzpatrick, P. J., J. A. Knaff, C. W. Landsea, and S. V. Finley, 1995: Documentation of a systematic bias in the aviation model's forecast of the Atlantic tropical upper-tropospheric trough: Implications for tropical cyclone forecasting. *Wea. Forecasting*, **10**, 433–446, https://doi.org/10.1175/1520-0434(1995)010<0433:DOASBI>2.0.CO;2.

Gamache, J. F., J. S. Griffin Jr., P. P. Dodge, and N. F. Griffin, 2004: Automatic Doppler analysis of three-dimensional wind fields in hurricane eyewalls. *26th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 5D.4, http://ams.confex.com/ams/pdfpapers/75806.pdf.

Gopalakrishnan, S. G., F. Marks, J. A. Zhang, X. Zhang, J.-W. Bao, and V. Tallapragada, 2013: A study of the impacts of vertical diffusion on the structure and intensity of the tropical cyclones using the high-resolution HWRF system. *J. Atmos. Sci.*, **70**, 524–541, https://doi.org/10.1175/JAS-D-11-0340.1.

Han, J., and H.-L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Wea. Forecasting*, **26**, 520–533, https://doi.org/10.1175/WAF-D-10-05038.1.

Harr, P. A., M. S. Kalafsky, and R. L. Elsberry, 1996: Environmental conditions prior to formation of a midget tropical cyclone during TCM-93. *Mon. Wea. Rev.*, **124**, 1693–1710, https://doi.org/10.1175/1520-0493(1996)124<1693:ECPTFO>2.0.CO;2.

Harris, L. M., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon. Wea. Rev.*, **141**, 283–306, https://doi.org/10.1175/MWR-D-11-00201.1.

——, ——, and C. Tu, 2016: High-resolution climate simulations using GFDL HiRAM with a stretched global grid. *J. Climate*, **29**, 4293–4314, https://doi.org/10.1175/JCLI-D-15-0389.1.

Hazelton, A. T., and R. E. Hart, 2013: Hurricane eyewall slope as determined from airborne radar reflectivity data: Composites and case studies. *Wea. Forecasting*, **28**, 368–386, https://doi.org/10.1175/WAF-D-12-00037.1.

——, R. Rogers, and R. E. Hart, 2015: Shear-relative asymmetries in tropical cyclone eyewall slope. *Mon. Wea. Rev.*, **143**, 883–903, https://doi.org/10.1175/MWR-D-14-00122.1.

Houze, R. A., Jr., S. S. Chen, B. F. Smull, W.-C. Lee, and M. M. Bell, 2007: Hurricane intensity and eyewall replacement. *Science*, **315**, 1235–1239, https://doi.org/10.1126/science.1135650.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Kaplan, J., M. DeMaria, and J. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **25**, 220–241, https://doi.org/10.1175/2009WAF2222280.1.

Knaff, J. A., M. DeMaria, B. Sampson, and J. M. Gross, 2003: Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92, https://doi.org/10.1175/1520-0434(2003)018<0080:SDTCIF>2.0.CO;2.

Landsea, C., J. Franklin, and J. Beven, 2015: The revised Atlantic hurricane database (HURDAT2). National Hurricane Center Rep., 6 pp., http://www.nhc.noaa.gov/data/hurdat/hurdat2-format-atlantic.pdf.

Lin, S.-J., 1997: A finite-volume integration method for computing pressure gradient force in general vertical coordinates. *Quart. J. Roy. Meteor. Soc.*, **123**, 1749–1762, https://10.1002/qj.49712354214.

——, 2004: A ''vertically Lagrangian'' finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2.

——, and R. B. Rood, 1997: An explicit flux-form semi-Lagrangian shallow-water model on the sphere. *Quart. J. Roy. Meteor. Soc.*, **123**, 2477–2498, https://doi.org/10.1002/qj.49712354416.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2.d

Lorsolo, S., J. Gamache, and A. Askoy, 2013: Evaluation of the Hurricane Research Division Doppler radar analysis software using synthetic data. *J. Atmos. Oceanic Technol.*, **30**, 1055–1071, https://doi.org/10.1175/JTECH-D-12-00161.1.

Mallen, K. J., M. T. Montgomery, and B. Wang, 2005: Reexamining the near-core radial structure of the tropical cyclone primary circulation: Implications for vortex resiliency. *J. Atmos. Sci.*, **62**, 408–425, https://doi.org/10.1175/JAS-3377.1.

Marks, F., K. J. Sellwood, and S. Abarca, 2016: Evaluation of the HWRF radius of maximum wind using Doppler radar analysis. *32nd Conf. on Hurricanes and Tropical Meteorology*, San Juan, PR, Amer. Meteor. Soc., https://ams.confex.com/ams/32Hurr/webprogram/Paper293852.html.

Nolan, D. S., J. A. Zhang, and D. P. Stern, 2009a: Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of in situ observations and high-resolutions simulations of Hurricane Isabel (2003). Part I: Initialization, maximum winds, and the outer-core boundary layer. *Mon. Wea. Rev.*, **137**, 3651–3674, https://doi.org/10.1175/2009MWR2785.1.

——, D. P. Stern, and J. A. Zhang, 2009b: Evaluation of planetary boundary layer parameterizations in tropical cyclones by comparison of in situ observations and high-resolution simulations of Hurricane Isabel (2003). Part II: Inner-core boundary layer and eyewall structure. *Mon. Wea. Rev.*, **137**, 3675–3698, https://doi.org/10.1175/2009MWR2786.1.

Reasor, P. D., R. Rogers, and S. Lorsolo, 2013: Environmental flow impacts on tropical cyclone structure diagnosed from airborne Doppler radar composites. *Mon. Wea. Rev.*, **141**, 2949–2969, https://doi.org/10.1175/MWR-D-12-00334.1.

Rogers, R., S. Lorsolo, P. Reasor, J. Gamache, and F. Marks, 2012: Multiscale analysis of tropical cyclone kinematic structure from airborne Doppler radar composites. *Mon. Wea. Rev.*, **140**, 77–99, https://doi.org/10.1175/MWR-D-10-05075.1.

——, P. D. Reasor, and S. Lorsolo, 2013: Airborne Doppler observations of the inner-core structural differences between intensifying and steady-state tropical cyclones. *Mon. Wea. Rev.*, **141**, 2970–2991, https://doi.org/10.1175/MWR-D-12-00357.1.

——, ——, and J. A. Zhang, 2015: Multiscale structure and evolution of Hurricane Earl (2010) during rapid intensification. *Mon. Wea. Rev.*, **143**, 536–562, https://doi.org/10.1175/MWR-D-14-00175.1.

Sadler, J. C., 1976: A role of the tropical upper tropospheric trough in early season typhoon development. *Mon. Wea. Rev.*, **104**, 1266–1278, https://doi.org/10.1175/1520-0493(1976)104<1266:AROTTU>2.0.CO;2.

Schmidt, F., 1977: Variable fine mesh in the spectral global models. *Beitr. Phys. Atmos.*, **50**, 211–217.

Stern, D. P., and D. S. Nolan, 2009: Reexamining the vertical structure of tangential winds in tropical cyclones: Observations and theory. *J. Atmos. Sci.*, **66**, 3579–3600, https://doi.org/10.1175/2009JAS2916.1.

——, and ——, 2011: On the vertical decay rate of the maximum tangential winds in tropical cyclones. *J. Atmos. Sci.*, **68**, 2073–2094, https://doi.org/10.1175/2011JAS3682.1.

——, J. R. Brisbois, and D. S. Nolan, 2014: An expanded dataset of hurricane eyewall sizes and slopes. *J. Atmos. Sci.*, **71**, 2747–2762, https://doi.org/10.1175/JAS-D-13-0302.1.

Stevenson, S. N., K. L. Corbosiero, and J. Molinari, 2014: The convective evolution and rapid intensification of Hurricane Earl (2010). *Mon. Wea. Rev.*, **142**, 4364–4380, https://doi.org/10.1175/MWR-D-14-00078.1.

Stewart, S. R., 2016: Tropical cyclone report Hurricane Danny (AL042015) 18–24 August 2015. National Hurricane Center Rep., 18 pp., http://www.nhc.noaa.gov/data/tcr/AL042015_Danny.pdf.

——, 2017: Tropical cyclone report Hurricane Matthew (AL142016) 28 September–9 October 2016. National Hurricane Center Rep., 96 pp., http://www.nhc.noaa.gov/data/tcr/AL142016_Matthew.pdf.

Velden, C. S., 1993: The relationship between tropical cyclone motion, intensity and the vertical extent of the environmental steering layer in the Atlantic basin. Preprints, *20th Conf. on Hurricanes and Tropical Meteorology*, San Antonio, TX, Amer. Meteor. Soc., 31–34.

Willoughby, H. E., 1990: Temporal changes of the primary circulation in tropical cyclones. *J. Atmos. Sci.*, **47**, 242–264, https://doi.org/10.1175/1520-0469(1990)047<0242:TCOTPC>2.0.CO;2.

——, J. A. Clos, and M. G. Shoreibah, 1982: Concentric eye walls, secondary wind maxima, and the evolution of the hurricane vortex. *J. Atmos. Sci.*, **39**, 395–411, https://doi.org/10.1175/1520-0469(1982)039<0395:CEWSWM>2.0.CO;2.

Zhang, J. A., S. Gopalakrishnan, F. D. Marks, R. F. Rogers, and V. Tallapragada, 2012: A developmental framework for improving hurricane model physical parameterizations using aircraft observations. *Trop. Cyclone Res. Rev.*, **1** (4), 1–11.

——, D. S. Nolan, R. F. Rogers, and V. Tallapragada, 2015: Evaluating the impact of improvements in the boundary layer parameterization on hurricane intensity and structure forecasts in HWRF. *Mon. Wea. Rev.*, **143**, 3136–3155, https://doi.org/10.1175/MWR-D-14-00339.1.

——, R. F. Rogers, and V. Tallapragada, 2017: Impact of parameterized boundary layer structure on tropical cyclone rapid intensification forecasts in HWRF. *Mon. Wea. Rev.*, **145**, 1413–1426, https://doi.org/10.1175/MWR-D-16-0129.1.

Zhao, Q., and F. H. Carr, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.*, **125**, 1931–1953, https://doi.org/10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2.

Zick, S. E., and C. J. Matyas, 2016: Evolving geometries in the precipitation patterns of 2004-2012 U.S. landfalling hurricanes. *Special Symp. on Hurricane Katrina*, New Orleans, LA, Amer. Meteor. Soc., 478, https://ams.confex.com/ams/96Annual/webprogram/Manuscript/Paper282702/zickAMS2016KatrinaExtendedvF.pdf.