Tropical Cyclone Intensity Estimation in the North Atlantic Basin Using an Improved Deviation Angle Variance Technique

ELIZABETH A. RITCHIE AND GENEVIEVE VALLIERE-KELLEY
Department of Atmospheric Sciences, The University of Arizona, Tucson, Arizona

MIGUEL F. PINEROS AND J. SCOTT TYO
College of Optical Sciences, The University of Arizona, Tucson, Arizona

(Manuscript received 22 December 2011, in final form 17 May 2012)

ABSTRACT
This paper describes results from an improvement to the objective deviation angle variance technique to estimate the intensity of tropical cyclones from satellite infrared imagery in the North Atlantic basin. The technique quantifies the level of organization of the infrared cloud signature of a tropical cyclone as an indirect measurement of its maximum wind speed. The major change described here is to use the National Hurricane Center’s best-track database to constrain the technique. Results are shown for the 2004–10 North Atlantic hurricane seasons and include an overall root-mean-square intensity error of 12.9 kt (6.6 m s⁻¹), where 1 kt = 0.514 m s⁻¹, and annual root-mean-square intensity errors ranging from 10.3 to 14.1 kt. A direct comparison between the previous version and the one reported here shows root-mean-square intensity error improvements in all years with a best improvement in 2009 from 17.9 to 10.6 kt and an overall improvement from 14.8 to 12.9 kt. In addition, samples from the 7-yr period are binned based on level of intensity and on the strength of environmental vertical wind shear as extracted from Statistical Hurricane Intensity Prediction Scheme (SHIPS) data. Preliminary results suggest that the deviation angle variance technique performs best at the weakest intensity categories of tropical storm through hurricane category 3, representing 90% of the samples, and then degrades in performance for hurricane categories 4 and 5. For environmental vertical wind shear, there is far less spread in the results with the technique performing better with increasing vertical wind shear.

1. Introduction
Tropical cyclone study and prediction has historically presented many obstacles. Forecasting tropical cyclone motion and landfall have improved greatly in the years between the mid-1970s and the turn of the century, in large part owing to the improved technology offered by satellite observations and considerable investment in the development of numerical weather prediction models (Ritchie et al. 2003). However, by comparison, intensity estimation and intensity forecasts still show only slowly improving skill (e.g., Fig. 1).

A major problem for both researchers and forecasters has been the lack of observations, particularly on the scales necessary to properly encapsulate tropical cyclone intensity processes. Most tropical cyclones develop over the ocean where there are few or no in situ surface or upper-air measurements. Whereas tropical cyclones can be detected by the Doppler radar network along the coastline, and in midocean by fleets of research aircraft, with radar, dropwindsondes, and many other remote and in situ instruments, the deployment of these aircraft is expensive. Furthermore, they may provide observations of intensity and intensity evolution in individual tropical cyclones; however, these observations do not provide a broad base from which to develop a deeper understanding of intensification. In addition, such observations rarely capture early intensification stages as these occur far from the U.S. mainland where these technologies are available.

There is another source of tropical cyclone observations that can be used for intensity estimation that is continuous and covers the entire tropical regions. Geostationary meteorological satellites have become a primary source of information about tropical cyclones due
to their global coverage and high temporal frequency. These instruments do not provide direct measurement of surface parameters such as wind speed, temperature, and humidity, but there are many algorithms in place that indirectly estimate such atmospheric variables from the image data. The current generation of geostationary satellites offers multispectral infrared (IR) measurements at 4-km resolution and visible spectrum measurements at 1-km resolution at intervals of 30 min or less. These images can be sequenced to create a time series of tropical cyclone cloud structure, which can be related back to the intensity of the tropical disturbance using a combination of empirical and theoretical models. These models relate the infrared brightness temperatures of the clouds in the satellite images to the deep convective activity associated with tropical cyclone development. In addition, the instruments on the geostationary satellites provide high temporal profiles of winds, temperature, and humidity in the atmosphere. However, none of these measurements include direct measurements of the maximum surface wind speed or minimum sea level pressure, which are the primary indicators of tropical cyclone intensity.

Due to the lack of direct in situ measurements of tropical cyclone intensity, several techniques have been developed to estimate the intensity based on indirect factors. Currently, the Dvorak technique is the most widely used technique for estimating tropical cyclone development via infrared satellite imagery (Dvorak 1975). The Dvorak technique uses a $T$-number scale to estimate tropical cyclone intensity based on cloud patterns of central and banding features in tropical cyclone infrared imagery (Velden et al. 1998). The combination of these features determines the overall shape and symmetry of the tropical cyclone, which are used to determine the current and near-future intensities of the tropical disturbance (Dvorak 1975). The earliest suggestion of a correlation between the most circularly symmetrical storms around a center point or eye and the most intense storms was determined and tested through the Dvorak technique. Other defining features that have been linked to a strong or intense–intensifying tropical cyclone are the brightness temperature; a solid, dense overcast at the center of the tropical cyclone with deep convection in the eyewall; and the spread of cirrus outward from the center. A weak or weakening cyclone will not display these features. Weakening can be caused by changes in the environment of the cyclone, including moving over land, increased vertical wind shear, increased dry air, or moving over colder waters (Dvorak 1975). In the application of the Dvorak technique, the aforementioned features that are linked with intensification, along with the determined $T$ number, are used to create an intensity curve, which can then be employed in forecasting the future of the tropical cyclone. However, the original Dvorak technique is very subjective, time intensive, and is dependent on the expertise and experience of the user. Despite this, the Dvorak technique is still used as the primary intensity forecasting tool in many tropical cyclone forecasting centers around the world (e.g., Velden et al. 1998, 2006b,c; Knaff et al. 2010).

The objective Dvorak technique (ODT) was introduced in the mid-1980s (Velden et al. 1998). The ODT uses computer-based algorithms to recognize patterns in cloud structure, identify environmental factors that might contribute to intensification or weakening, and create an intensity estimate based on a lookup table, which is determined by observation and practical experience. Although these innovations make the ODT less subjective (Velden et al. 1998), the center of the tropical cyclone must still be either manually located by an analyst or determined from external sources (Piñeros et al. 2010). Olander and Velden (2007) developed the advanced Dvorak technique (ADT), which introduces new procedures in making an intensity estimate from satellite-based imagery including the introduction of regression equations to estimate tropical cyclone intensity.

In addition, there are other techniques for estimating intensity based on satellite measurements in developmental
stages. Kossin et al. (2007) recently described a new satellite-based technique in which the radius of maximum wind, the critical wind radii, and the two-dimensional surface wind field are estimated using mean 12-h IR imagery. The technique estimates the two-dimensional wind field from the IR imagery and then validates the field against aircraft wind observations. Furthermore, techniques for estimating the intensity of a tropical cyclone have also been developed using measurements from the Advanced Microwave Sounding Unit (AMSU; Spencer and Braswell 2001; Demuth et al. 2004). Some of these techniques have been combined to enhance the tropical cyclone intensity estimation (e.g., Velden et al. 2006a).

A promising and recently developed method called the deviation angle variance (DAV) technique uses the gradient of the brightness temperature field to determine the level of symmetry of the tropical cyclone cloud structure (Piñeros et al. 2008, 2011). The departure of the cloud field from axisymmetry has been shown to be correlated with the intensity of the tropical cyclone. The basic technique has achieved seasonal root-mean-square (RMS) intensity error scores as low as 13–15 kt (6.7–7.7 m s⁻¹, where 1 kt = 0.514 m s⁻¹). However, a systematic problem with the technique for cases that are very weak with offset circular cloud signatures and an exposed low-level circulation (e.g., Tropical Storms Ana and Erika in 2009) has led to an important modification in DAV that is presented here. The technique described in Piñeros et al. (2011) used the minimum DAV value calculated anywhere within the cloud system to compute the intensity estimate. In most instances this location of minimum variance did not exactly coincide with the true tropical cyclone (TC) center, and in some instances the minimum variance location was far removed from the true storm center (see Figs. 5–7 in Piñeros et al. 2008). While the aim of the DAV development has been to create a completely objective tool, we have, to date, been unable to provide a completely robust center estimator. However, in many operational scenarios, information about TC centers is available, and this information could be fed into the DAV computation when it is available, much as it is done for several other subjective and objective intensity estimators currently in operational use. To provide a “best case” DAV intensity estimate, this paper uses a center fix from the National Hurricane Center’s (NHC) “best track” database to locate the position where the DAV will be calculated. This combination of using best-track information to inform the DAV technique yields even lower RMS intensity error scores (14% reduction in RMS error on average) than were previously achieved and improves the DAV technique to operational standards. While it is true that the best-track center would not be available to forecasters in real time, the injection of some expertise into the DAV system can eliminate the most egregious problem cases, like Tropical Storms Ana and Erika in 2009.

The paper is structured as follows. The data and methodology of the DAV technique are reviewed in section 2. Results are presented in section 3 and an analysis of the technique’s performance is provided in section 4. Finally, a summary, conclusions, and future directions to continue improving the technique are provided in section 5.

2. Methodology

a. Data

This study focuses on Atlantic Ocean tropical cyclones during the 7-yr period 2004–10 (Franklin et al. 2006; Beven et al. 2008; Franklin and Brown 2008; Brennan et al. 2009; Brown et al. 2010; Berg and Avila 2011). Those times when a tropical cyclone center passed over continents and large islands were removed from the training set. Observations show that tropical cyclones that make landfall rapidly decay at a rate that is different than dissipation rates for over ocean tropical cyclones and a different set of parametric curves, still under development, are required for these cases. In addition, those storms that spend the majority of their life cycle outside of the area of the rectified satellite imagery are also removed from the database. This left a total 95 tropical cyclones in the dataset.

The data used in the study are the Geostationary Operational Environmental Satellite (GOES-E) longwave (10.7 μm) infrared (IR) images with a 4-km nadir resolution, captured at 30-min intervals. Seven years of data from the North Atlantic tropical cyclone season, which officially covers the period 1 June through 30 November, were collected. These images were rectified from a natural earth coordinate system to a Mercator projection, removing much of the differences in pixel resolution due to the earth’s curvature. The final images were resampled to have uniform spatial resolution of 10 km per pixel, centered at 20°N and 67°W and cover an area from 3.5° to 36.5°N and 105.1° to 28.9°W.

Tropical cyclone best-track data from the NHC were obtained online for the period of interest (http://www.nhc.noaa.gov/pastall.shtml). These data are archived at 6-h intervals. Thus, to match center locations and wind speed estimates to the 30-min temporal resolution of the satellite data, the best-track data were linearly interpolated to match the satellite temporal resolution. The interpolated best-track center latitude-longitude locations were then converted into pixel values using the approximation that 1 pixel is equal to 10 km. Therefore, in the longitudinal direction, 1° is equal to 111.325 km = 11.32567 km.
pixels and in the latitudinal direction $1^\circ$ is equal to $\sec(\theta) \times 111.325$ km = $\sec(\theta) \times 11.325$ pixels, where $\theta$ is degrees latitude. A visual check ensured that the conversion to pixel values was correctly performed. Note that there can occasionally be a small error in the pixel center compared with the IR center because best-track center positions are interpolated linearly to 30-min resolution resulting in some error between 6-h synoptic times that were manually corrected. A subsequent implementation of this technique uses spline interpolation, which better captures the actual track of the TC in between synoptic best-track fixes, and requires less manual intervention, but this automatic smoothing was not implemented for this study.

b. Method

The estimation of tropical cyclone intensity is a three-step procedure described in Piñeros et al. (2008, 2011) and is briefly summarized here (Fig. 2). First, a low-pass filter is applied to the image to remove noise, then the gradient of the brightness temperatures in the image is calculated from the $x$ and $y$ derivatives by using Sobel’s template (Gonzalez and Woods 2002; Bow 2002, 311–312) (Figs. 2a,d). Next, a radial line is extended from the pixel at the center of the tropical cyclone to every other pixel within a defined radius of the center (Figs. 2b,e). The angle between the direction of the gradient vector at each pixel and the radial is then calculated (the deviation angle) and the probability distribution function of the deviation angle is estimated. Representative histograms are plotted in Figs. 2c,f. If the tropical cyclone is very close to a perfectly axisymmetric vortex, then all the gradient vectors will point along the radial lines to the center of the tropical cyclone and the deviation angles will be nearly zero (Figs. 2a–c). Finally, the variance of the deviation angle is used as a measurement of the “organization” of the tropical cyclone. For an intense axisymmetric tropical cyclone with highly organized cloud structure, this variance will be near zero, while for a weak disorganized tropical depression this variance will be high because the cloud structure, and therefore the brightness gradient, is extremely disorganized. Thus, the higher the “deviation angle variance” or DAV, the less intense the storm will be (Piñeros et al. 2008). Because the DAV varies depending on the area over which the calculation is made, the DAV in this study was calculated for 8 different radii from the center; 150 km through 500 km, every 50 km, resulting in eight DAV values for every image. Figure 3 shows the negative relationship between DAV and wind speed intensity for all eight radii of calculation for a case of Hurricane Paula (2010).

The major difference between the DAV calculated in Piñeros et al. (2011) and here is that in Piñeros et al. (2011) the DAV is calculated for every pixel in the satellite scene. Then, the minimum DAV value is extracted for every tropical cyclone in the scene, regardless of where in the cloud cluster it is located, and correlated to
intensity. In the modification described here, the DAV at the pixel location that corresponds to the NHC best-track center location is the value that is correlated to intensity. Furthermore, to allow for the possibility of a slight error in the center position, the final DAV value is calculated as an average over the nine pixels surrounding the NHC best-track center point.

Once the DAV statistics are computed for each storm, a training set comprising all the tropical cyclones with wind speeds greater than or equal to 34 kt from the 7 yr from 2004 to 2010 are compared to the corresponding interpolated best-track intensity estimates. Results are similar if winds speeds less than 34 kt are incorporated into the training and testing, but because of a lack of validating aircraft reconnaissance observations at these lower wind speeds, we currently do not incorporate them in our statistics. A scatterplot of DAV to best-track intensity values is created and a sigmoid curve is fitted to the data for each radii of DAV calculation to describe the relationship between the best-track intensity estimates and the DAV values (Fig. 4). The sigmoid equation used for the variance–intensity relationship is given by

\[
P(t) = \frac{140}{1 + e^{a(x-b)}} + 25,
\]

where \(x\) is the calculated DAV, and \(a\) and \(b\) are the parameters that are determined using the training data. They represent the slope and shift, respectively, of the sigmoid curve fitted to the scatterplot. The lower sigmoid limit is set to 25 kt and 165 (140 + 25) kt is the upper limit for the sigmoid. These values were established by experimentation and act to limit the possible intensity values that the DAV technique could estimate. One of the advantages of the sigmoid over other possible curves such as polynomials is that the ends of the curve are bounded by the observations such that unrealistically high and low estimates of intensity are not possible. A fifth-order polynomial was also tested, but produced higher overall RMS intensity errors. The sigmoid relationships obtained using each of the eight radii of DAV calculations are used for the results in this paper.

The final step in the procedure is to produce actual estimates of the intensity of an independent set of tropical cyclones using the DAV–intensity relationship obtained by training on 6 of the 7 yr from 2004 to 2010 and testing on the seventh year. In this manner, results are obtained for 7 yr of testing tropical cyclones. Although this exact
Table 1. RMSE (kt) for every radius of DAV calculation (150–400 km), number of storms, and number of samples with best-track estimates ≥33 kt, for each individual year. The lowest annual RMSE errors are highlighted in boldface italics.

<table>
<thead>
<tr>
<th>Year</th>
<th>150 km</th>
<th>200 km</th>
<th>250 km</th>
<th>300 km</th>
<th>350 km</th>
<th>400 km</th>
<th>Storms</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>14.0</td>
<td><strong>13.3</strong></td>
<td>13.7</td>
<td>15.4</td>
<td>16.6</td>
<td>17.3</td>
<td>13</td>
<td>3122</td>
</tr>
<tr>
<td>2005</td>
<td>14.7</td>
<td><strong>14.1</strong></td>
<td>14.4</td>
<td>15.4</td>
<td>16.5</td>
<td>17.2</td>
<td>22</td>
<td>3793</td>
</tr>
<tr>
<td>2006</td>
<td>12.1</td>
<td>10.5</td>
<td><strong>10.3</strong></td>
<td>11.2</td>
<td>12.5</td>
<td>14.4</td>
<td>9</td>
<td>1858</td>
</tr>
<tr>
<td>2007</td>
<td><strong>11.4</strong></td>
<td>13.4</td>
<td>16.2</td>
<td>18.1</td>
<td>18.7</td>
<td>19.3</td>
<td>9</td>
<td>1232</td>
</tr>
<tr>
<td>2008</td>
<td>15.5</td>
<td>12.6</td>
<td><strong>12.0</strong></td>
<td>12.8</td>
<td>14.0</td>
<td>15.3</td>
<td>15</td>
<td>2925</td>
</tr>
<tr>
<td>2009</td>
<td>15.1</td>
<td>13.4</td>
<td>12.4</td>
<td>11.1</td>
<td><strong>10.6</strong></td>
<td><strong>10.6</strong></td>
<td>8</td>
<td>1095</td>
</tr>
<tr>
<td>2010</td>
<td>14.5</td>
<td>12.4</td>
<td><strong>11.8</strong></td>
<td>12.3</td>
<td>12.8</td>
<td>13.2</td>
<td>19</td>
<td>3598</td>
</tr>
<tr>
<td>All years</td>
<td><strong>12.8</strong></td>
<td><strong>12.9</strong></td>
<td>13.8</td>
<td>14.7</td>
<td>15.4</td>
<td>95</td>
<td>17623</td>
<td></td>
</tr>
</tbody>
</table>

3. Results

a. Relationship between DAV intensity and best-track wind speed

Previous work has shown that deviation angle variance is negatively correlated with intensity as measured by the maximum sustained surface wind speed. That is, as the deviation angle variance decreases, intensity increases (Piñeros et al. 2008). This same work has also shown that the DAV intensity estimates contain high-frequency oscillations (i.e., higher than the diurnal and semidiurnal time frames) due to the convective nature of the cloud systems that are being processed. Thus, the DAV signal is smoothed before conversion to intensity using a low-pass filter (impulse response: $e^{-\tau}$ with a filter time constant $\tau$ of 100 h). This affects the DAV signal in two ways: 1) it increases the correlation between the DAV-estimated intensity and the interpolated best-track estimates allowing for a better comparison and 2) it introduces a time lag into the signal, which may decrease the technique’s performance in rapidly intensifying cases, as was discussed elsewhere (Piñeros et al. 2008, 2011). The test year intensity estimates were compared with the actual interpolated intensities from the NHC best-track dataset by calculating the RMS intensity error between the two intensity signals. Average RMS intensity errors for each test year using several different radii of DAV calculation are shown in Table 1.

According to Table 1, RMS intensity errors vary between 10.3 and 14.1 kt for the individual years. Tweaking the sigmoid curve changes these values slightly, but not in any consistent way. In addition, the technique performed better for different radii in different years. In general, all but two years (2007 and 2009) had a best training radius of 200–250 km. This suggests that in the North Atlantic basin, the axisymmetric structure of the clouds within about 200–250 km of the center as measured by the DAV technique has the highest correlation with intensity of the tropical cyclone. It is not clear at this time why the best radii for the DAV calculation were higher (350–400 km) for 2009.

b. Comparison with the Piñeros et al. (2011) technique

Table 2 shows a direct comparison of the RMS errors between the two versions of the technique for all years. In the original technique, a high RMS intensity error of 17.9 kt was achieved in 2009 and analysis showed that this was due in part to two particular tropical cyclones that were very poorly estimated by the technique (Fig. 5):

Table 2. A comparison of RMS intensity errors for all years (2004–10) using the Piñeros et al. (2011) technique and the new center-based DAV method described here.

<table>
<thead>
<tr>
<th></th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (kt)</td>
<td>15.6</td>
<td>17.3</td>
<td>11.7</td>
<td>12.8</td>
<td>12.2</td>
<td>17.9</td>
<td>12.3</td>
<td>14.8</td>
</tr>
<tr>
<td>Best radius (km)</td>
<td>250</td>
<td>250</td>
<td>300</td>
<td>150</td>
<td>300</td>
<td>200</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>Center-based technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE (kt)</td>
<td><strong>13.3</strong></td>
<td>14.1</td>
<td>10.3</td>
<td>11.4</td>
<td>12.0</td>
<td>10.6</td>
<td>11.8</td>
<td>12.9</td>
</tr>
<tr>
<td>Best radius (km)</td>
<td>200</td>
<td>200</td>
<td>250</td>
<td>150</td>
<td>250</td>
<td>350</td>
<td>250</td>
<td>250</td>
</tr>
</tbody>
</table>
Tropical Storm (TS) Ana, with an average RMS intensity error of 25.5 kt, and TS Erika, with an average RMS intensity error of 27.4 kt (Table 2). Both of these storms were weak tropical storm–strength systems with maximum intensities between 35 (Ana) and 45 (Erika) kt, which, as is discussed in the next section, lends them to being overestimated by the technique because they lie at the lower limit of the sigmoid curve. In addition, both storms were located in moderate to high average environmental 200–850-hPa vertical wind shear (11–15 kt) for much of their lives. As a result, both storms were sheared storms with very circular uniform cloud shields and exposed low-level centers (Fig. 6). The use of the best-track center to pinpoint the DAV estimates was designed to improve cases similar to Ana and Erika. In these two cases, the best-track center modification improved the intensity estimation for Ana to 19.1 kt and for Erika to 6.0 kt. Two additional storms in 2009 were also embedded in moderate to high shear for much of their lifetime and showed significant improvement with the use of the new technique: Bill from 16.0 to 7.4 kt and Ida from 18.7 to 7.3 kt.

In general, the overall RMS intensity error is improved for all years using the center-determined DAV technique (Table 3), although the dramatic improvement in 2009 is not repeated in any other year. Overall, it can be concluded that the technique has been improved from the Piñeros et al. (2011) version in the Atlantic basin by using the best-track center-determined DAV intensity estimate. This is the version of the technique now adopted for use in other basins (e.g., M. F. Piñeros et al. 2012, unpublished manuscript) and also being developed for operational assessment at the National Hurricane and Joint Typhoon Warning Centers.

c. Results binned by tropical cyclone intensity

The relationship between TC intensity and DAV is reexamined here based on levels of intensity in order to ascertain whether the DAV performs better for different intensity bins. Each sample through the 7-yr period

FIG. 5. Time series of NHC best-track intensity estimates and the DAV intensity estimates for the 2009 Atlantic season. The two poorly estimated storms (Ana and Erika, which originally had individual RMS intensity errors of 20.7 and 26.8 kt, respectively) are shaded.

FIG. 6. Satellite infrared images of TSs Ana and Erika at (a) 0615 UTC 12 August and (b) 2045 UTC 1 September 2009, respectively, demonstrating the circular, but offset, central dense overcasts that were typical of these systems. The low-level center location is indicated by an asterisk in (a) and a star in (b).
was grouped into one of six bins based on the Saffir–Simpson categorization using the best-track intensity value: tropical storm (<64 kt) and hurricane categories 1 (64–82 kt), 2 (83–95 kt), 3 (96–113 kt), 4 (114–135 kt), and 5 (>135 kt). The RMS intensity error was then calculated for each group to ascertain if there were particular intensity bins for which the DAV technique performed better or worse than others. Table 3 shows the results of grouping by these intensity bins. In general there was a systematic increase in error as the intensity values increased. The tropical storm category had the highest number of samples (9896) and an RMS intensity error equal to 11 kt (Fig. 7). Figure 7 and Table 3 both suggest that the error is more heavily due to an overestimation of the wind speed for this group. Approximately 57% of the samples (5657) were overestimated by some amount by the technique. This bias toward overestimation at the tropical storm level of intensity is likely because of the bounding of the sigmoid curve relationship at that low wind speed end that results in a general overestimation of the weaker systems.

The RMS intensity errors for all samples falling in hurricane categories 1–3 were 12.5, 12.5, and 12.6 kt, respectively, with a total of 5862 samples. Figure 8 shows the combined estimates for categories 1–3. Both Fig. 8 and Table 3 suggest that the error is distributed evenly between over- and underestimation of the wind speed at the categories 1 and 2 levels, but tends toward underestimation for category 3 (Table 3).

The category 4 hurricanes had a total 1513 samples and an overall RMS intensity error equal to 17.7 kt (Table 3). Again, these storms were systematically underestimated with 80% of the sample being underestimated by some amount (Table 3). The category 5

<table>
<thead>
<tr>
<th>Bin</th>
<th>No. of samples</th>
<th>RMSE (kt)</th>
<th>No. overestimated (%)</th>
<th>No. underestimated (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tropical storms</td>
<td>9896</td>
<td>11.0</td>
<td>5657 (57)</td>
<td>4239 (43)</td>
</tr>
<tr>
<td>Hurricane category 1</td>
<td>2892</td>
<td>12.5</td>
<td>1620 (56)</td>
<td>1272 (44)</td>
</tr>
<tr>
<td>Hurricane category 2</td>
<td>1522</td>
<td>12.5</td>
<td>676 (44)</td>
<td>845 (56)</td>
</tr>
<tr>
<td>Hurricane category 3</td>
<td>1453</td>
<td>12.6</td>
<td>532 (37)</td>
<td>920 (63)</td>
</tr>
<tr>
<td>Hurricane category 4</td>
<td>1513</td>
<td>17.7</td>
<td>309 (20)</td>
<td>1204 (80)</td>
</tr>
<tr>
<td>Hurricane category 5</td>
<td>347</td>
<td>32.4</td>
<td>0 (0)</td>
<td>347 (100)</td>
</tr>
</tbody>
</table>

**TABLE 3.** The number of samples, RMSE (kt), and the number of samples (and %) either overestimated or underestimated for all samples categorized into bins based on the Saffir–Simpson scale of tropical cyclone intensities.

![Fig. 7](image-url)

**Fig. 7.** Intensity estimates calculated using the DAV technique compared with the best-track intensity estimates for all samples in the TS category. The samples are split evenly between (a) and (b) and do not represent a time-continuous plot.
hurricanes had the smallest number of samples (347) and an overall RMS intensity error equal to 32.4 kt (Table 3). The error with these samples is clearly due to the overall underestimation by the technique, as highlighted in Fig. 9, which shows the combined estimates of storms in categories 4 and 5 (see also Table 3), and is likely because of two main factors: the first being the lack of samples at the highest winds speeds in the training set, which results in a general underestimation of these highest wind speeds.

However, the second factor accounts for the general trend that moves from overestimation at the tropical storm intensities, to an even distribution for category 1 and 2 hurricanes, to increasing underestimation as the hurricane intensity categories increase through category 5 on the Saffir–Simpson scale. Figure 4 shows the sigmoid curve that is fitted to all samples using the entire period 2004–10 as the training set. This curve is representative of any of the training curves used to test the individual years (Table 1). It can be seen that the bounding of the two ends of the sigmoid curve leads to a systematic overestimation at the low end in Fig. 4 and by the time the curve has passed through 75 kt, it is underestimating the distribution. It is clear that further investigation into the most appropriate curve to represent the DAV–wind speed relationship is warranted, and significant improvement at the higher intensities is likely using this technique if a better fit to the training data can be found. Testing has already been done by varying some of the parameters that are passed into the sigmoid curve, and also by using a fifth-order polynomial fit to the data instead of a sigmoid fit. In the former case, the sigmoid curve used in this paper represents the best results using sigmoid curves, and there was no improvement to the estimates using the fifth-order polynomial curve.

One other factor in particular stands out at these highest intensities, especially since the majority of category 5 hurricanes occurred during the extended 2005 hurricane season. Examination of Fig. 10 suggests that sometimes rapid intensification is not handled well by the DAV technique. Hurricane Rita is a case where the DAV technique both lagged the onset of rapid intensification and also did not deepen the TC enough. This lag in the intensity estimation is at least partly due to the time filter that is applied to the DAV signal to smooth some of the high-frequency components and allow a better comparison and fit to the 6-h best-track intensity values. Some testing of different smoothing functions was done during the development of the estimation technique, and the current filter is the best option found thus far. This is an area of possible improvement to the intensity estimation technique and so is still a subject of future work.

The smoothed bias of the DAV intensity estimates compared with the NHC best-track intensity estimates is
plotted in Fig. 11, which summarizes the increasing negative bias in the DAV intensity estimates as intensity increases. An early effort to factor this bias into the estimates has met with only limited success. While some small improvement in the overall RMSE at the category 4 and 5 intensities has been achieved (1–3 kt), degradation of the intensity estimates at the lower intensities more than offsets any improvement in the overall statistic. Further work exploring ways to factor the bias into the estimation and to limit the bias all together is being pursued.

In summary, it can be concluded that the DAV technique has been improved over the Piñeros et al. (2011) version by fixing the technique to the best-track center position. In addition, the technique estimates tropical cyclone intensity most accurately at intensities of tropical storm through category 3 hurricane strength, which constitutes almost 90% of the total samples. Low-intensity storms in the tropical storm category are generally overestimated, as evidenced in Table 3. Hurricane categories 4 and 5 presented the greatest difficulty for the technique due to consistent underestimation of high-intensity samples (Table 3).

4. Analysis and discussion

Although the DAV technique performs well overall, and results are considerably improved from the previous version by using a predetermined center point, the performance of the technique has some dependence on storm intensity (Table 3). As highlighted in sections 3a and 3b, the periods where the DAV did not perform so well tend to fall into three main categories: overestimation (underestimation) at the low (high) intensities due to the choice of bounds on the sigmoid curve; difficulties for individual weak cases during periods of moderate to high vertical wind shear despite corrections using the operational center fix, suggesting that a different DAV–intensity relationship may be required during these periods; and, for 2005 in particular, difficulty with periods of rapid intensification because of the intrinsic lag in the DAV technique due to the particular low-pass filter employed here.

a. Technical limitations of the technique

During the original development of the DAV technique the bounds for the sigmoid were tested and set to reflect the most efficient means for establishing realistic limits on both the high and low ends of the wind speed and variance values (Piñeros et al. 2011). These bounds were determined by examining the range of high and low DAVs versus best-track intensity values and were set, in particular, to ensure that unrealistically high values of intensity could not be estimated by the technique. Inspection of Fig. 4 indicates that an exponential curve might “fit” the data as well, if not better, than the...
sigmoid curve shown in the figure. However, the exponential curve would have resulted in unrealistically high values of intensity for the DAV values being calculated (e.g., a DAV of 400^2 would give an intensity estimate of about 175 kt with an exponential fit). The current settings for the sigmoid bounds could be reexamined as more data points become available for training in order to better characterize the upper bound of the DAV–intensity relationship. The high RMS intensity error for all samples greater than category 4 hurricanes highlights this issue. A similar problem occurs at the low end of the intensity curve where there is a large range of possible variance values for the intensity range of 35–50 kt and the curve fit is perhaps on the high side. However, the overestimation that occurs in the tropical storm category is relatively small, as evidenced by an overall RMS intensity error of 10.7 kt despite the category comprising over half of the testing samples (56%).

1) WEAK, TROPICAL STORM INTENSITY CASES

As discussed in section 3c, the weak, tropical storm category samples are consistently overestimated by the DAV technique. The bias for the intensity range 35–40 kt is between +12 and 0 kt. While this overestimation can be attributed to the sigmoid fit of the data at this intensity range, there are some other possible factors that may come into play and can be best illustrated by individual examples.

One particular very weak storm that was overestimated by the DAV technique due in part to the lower sigmoid bound was the previously mentioned TS Ana from the 2009 season, with an overall DAV intensity error of 19.1 kt (Fig. 12). TS Ana spent its entire life over the open ocean and thus the entire inventory of best-track intensity estimates of Ana are based on the Tropical Analysis Forecast Branch satellite Dvorak estimates of intensity (Berg and Avila 2011). TS Ana never exceeded a maximum sustained wind speed of 35 kt, putting it squarely in the overestimation end of the bias curve. The only verifying aircraft reconnaissance observations were measured at the very end of its life cycle after it had dissipated. Of note, Ana was also a sheared storm with an average shear value from SHIPS data (DeMaria and Kaplan 1994) calculated between 850 and 200 hPa for the lifetime of the storm to be >11 kt. As a result, the storm produced a very circular, uniform, but offset cloud pattern for much of its life (e.g., Fig. 12a) and therefore would be more likely to cause the DAV to overestimate the intensity. The current implementation of the DAV technique using best-track center locations to pinpoint the DAV value helped to offset this problem, reducing TS Ana’s overall RMSE from 25.5 to 19 kt. In spite of this correction, the DAV intensity estimate for Ana remained relatively high.

2) INTENSE, HURRICANE CATEGORY 4–5 CASES

As discussed in section 3c, the intense category 4 and 5 hurricanes are consistently underestimated by the DAV technique. The bias for the intensity range 134 kt is between −20 and −70 kt (Fig. 11). This underestimation can be almost entirely attributed to the sigmoid fit of the data at this intensity range because of a lack of samples. There is another factor that does come into play during periods of rapid changes in intensity that is well illustrated by Hurricane Rita (2005).

Rita reached a maximum sustained surface wind speed of 155 kt, a category 5 hurricane, with minimum sea level pressure of 897 hPa at 0600 UTC 22 September,
before weakening prior to making landfall near the Texas–Louisiana border. Figure 13 shows an image of Rita near peak intensity in the left panel along with the best-track and DAV intensity traces in the right panel. It can be seen that Rita is considerably underestimated by the DAV technique at the highest intensities. Note that immediately after peak intensity Rita rapidly weakens from 155 to 125 kt in just 12 h. As the DAV intensity estimates approach their peak values (near 140 kt in this case), the filter that was applied to the DAV signal would already be incorporating these rapidly decreasing values into the smoothing, resulting in a peak estimate much lower than that actually realized. That is, because of the combination of the way the filter is applied, and Rita’s rapid climb and fall about its peak value, it would not be possible for the DAV to reach the actual peak intensity value in this case.

In addition to the inability of the technique to capture Rita’s highest intensities, the DAV lags the period of more rapid intensification by ~10 h, which is likely a function of the low-pass filter that is used to smooth the DAV signal. The use of 30-min data for the DAV signal results in high-frequency oscillations that are not present in the 6-h best-track data even when interpolated to 30-min resolution. The low-pass filter is used to smooth out these oscillations allowing for a better correspondence with the best-track intensity values. However, this does introduce a lag into the signal, which, in the case of Hurricane Rita, translates as a delay in the rapid intensification period, although the actual rate of intensification is fairly well represented. In Fig. 13, the unfiltered DAV intensity estimate time series is also overlaid with the gray solid line. The high-frequency oscillations that are smoothed by the low-pass filter can be easily observed in this time series. In addition, it can also be seen (using the human eye) that the unfiltered time series not only captures the timing of the rapid intensification phase, but also does a much better job of estimating the peak intensity of Rita as well as the subsequent weakening. We are in the process of examining ways to smooth the data in such a way that the high correlation between the DAV signal and best-track intensity is maintained but the induced lag and lowering of peak intensity estimates are mitigated.

b. Consistent trends associated with environmental vertical wind shear

Vertically sheared environments are generally unfavorable for tropical cyclones because they disrupt the rising of moist, warm air for developing cases and interfere with the overall organization in developed cases. If strong enough, environmental vertical wind shear may ultimately weaken or destroy the tropical cyclone. Tropical Storms Ana and Erika from 2009, which were discussed briefly in section 3b, are examples of tropical cyclones that were embedded in moderate to strong environmental vertical wind shear for much of their lives. The changes in the DAV technique described in this paper were specifically designed to improve these types of cases. The samples were binned by vertical wind shear increments similar to the intensity bins described in section 3c to examine whether environmental vertical wind shear biases the DAV intensity results. The 200–850-hPa environmental vertical wind shear was extracted from SHIPS data (DeMaria and Kaplan 1994) for all samples from 2004 to 2010 and the results are shown in Table 4. The results in Table 4 suggest that there is some dependence of DAV intensity estimate errors on the strength of the environmental vertical wind shear, although the signal is much less than that for intensity categories. The range of RMS

FIG. 13. Example of Hurricane Rita from the 2005 season: (a) IR satellite image valid for 1800 UTC 21 Sep, intensity of 145 kt, and pressure of 920 hPa. The small eye is easily observed in the IR image. (b) Best-track (solid line), smoothed DAV (dashed line) intensity estimate, and unfiltered DAV intensity estimate (gray solid line) for Hurricane Rita. Overall RMSE = 14.3 kt.
intensity errors for vertical wind shear values is 9.5–14.2 kt (Table 4) with the RMS intensity error values increasing for decreasing environmental vertical wind shear. The extremely low RMS intensity error for very high (>$50$ kt) shear values is simply a reflection of the low number of samples (two) in that category. The correlation coefficient between the DAV intensity estimate errors and SHIPS environmental vertical wind shear is 0.1. The bias of the DAV intensity estimates compared with the SHIPS environmental vertical wind shear is plotted in Fig. 14a along with a scatterplot of the difference between the best-track and DAV intensity estimates versus SHIPS-based vertical wind shear values (Fig. 14b). The bias indicates a slight tendency toward underestimation at very low vertical wind shear values, which is almost completely explained by the tendency for underestimation at the high-intensity (and thus low vertical wind shear) end of the DAV curve. There is also a slight overestimation at mid- to-high vertical wind shear values, which is again most likely explained by the tendency for overestimation at the low-intensity (and generally higher vertical wind shear) end of the DAV curve. However, the scatterplot illustrates the near-zero correlation between the two variables, and it is difficult to extract any real trends associated with the magnitude of vertical wind shear and intensity. Based on these results, it is tentatively concluded that any bias in the DAV technique is minimal at high vertical wind shear.

5. Conclusions

An improvement to the deviation angle variance technique first discussed in Piñeros et al. (2008) and expanded in Piñeros et al. (2011) is described and evaluated. The basic technique characterizes the development of tropical cyclones based on the departure of their cloud structure from axisymmetry. In the version of the technique described and assessed in this paper, the NHC best-track center location of each tropical cyclone is used to pinpoint where the axisymmetry parameter is calculated. Overall, the refined DAV technique is shown to produce less error than previous versions of the DAV technique developed in Piñeros et al. (2008, 2011) with the best annual RMS intensity errors as low as 10.3 kt and an improvement of the worst year (2009) reported in Piñeros et al. (2011) from 18 to 10.6 kt.

Results for samples categorized by intensity bins show that the refined DAV technique performs best at intensities of tropical storms and hurricane categories 1–3 with RMS intensity errors of 11.0 and 12.5 kt, respectively. While these lower-intensity categories compose the majority (~90%) of the samples in the 7-yr dataset, the most intense samples (hurricane categories 4 and 5) clearly pose a problem for the technique with an overall combined error of 21.2 kt. The main problems in the DAV technique identified here include the lower and upper bounds of the sigmoid curve used to fit the training data, a lack of training samples at the highest intensity values, and the smoothing function used in the DAV technique that introduces a lag in the signal and a lowering of peak

![Fig. 14.](image-url)
intensities. Environmental vertical wind shear was not identified as a significant issue for the technique since no strong systematic biases in the RMS intensity error could be identified as vertical wind shear increased.

In general, the results of this research continue to support the basic DAV premise that more axisymmetric storms produce higher-intensity values, as well as reinforcing the ability of the DAV technique to quantify and analyze the level of axisymmetry of a tropical cyclone for use in intensity estimation. While the DAV technique holds great promise as a tool for real-time analysis of tropical cyclone intensity, there are still various elements of the DAV technique that could be improved with further analysis and modification. Future directions for this research include further analysis into the bounding of the training sigmoid to determine more appropriate limits, which will improve the over- and underestimation that occurs at the lower and upper ends of the intensity spectrum, respectively. In addition, methodologies to improve the technique include using only periods when aircraft reconnaissance observations are available for training so that the best possible intensity estimates are employed to train the DAV technique, mapping multiple DAV curves for different intensity bins, and mapping multiple DAV curves for different Dvorak cloud scene types. Finally, the error statistics based on the same intensity bins and Dvorak cloud scene mappings will be documented to improve the overall operational utility of the DAV technique.

Acknowledgments. Best-track tropical cyclone files were downloaded from the National Hurricane Center website (http://www.nhc.noaa.gov/pastall.shtml). SHIPS data were obtained online (http://rammb.cira.colostate.edu/research/tropical_cyclones/ships/developmental_data.asp). Special thanks are given to Dr. Yolande Serra for her constructive suggestions during the latter part of the study and to the comments provided by two anonymous reviewers. This project was supported by the National Oceanographic Partnership Program (NOPP) under Grant N00014-10-1-0146.

REFERENCES


