

An Analysis of NCEP Tropical Cyclone Vitals and Potential Effects on Forecasting Models

SAM TRAHAN

*NOAA/Environmental Modeling Center, Camp Springs, and University of Maryland,
Baltimore County, Baltimore, Maryland*

LYNN SPARLING

University of Maryland, Baltimore County, Baltimore, Maryland

(Manuscript received 30 May 2011, in final form 21 October 2011)

ABSTRACT

This study analyzes the Tropical Cyclone Vitals Database (TCVitals), which contains cyclone location, intensity, and structure information, generated in real time by forecasters. These data are used to initialize cyclones in several NCEP operational forecasting models via bogusing and vortex relocation methods. In many situations, time is of the essence and the TCVitals database represents the best real-time estimate of the cyclone state possible in real time, given the limitations of available data and time constraints inherent in real-time forecasting. NCEP and other users of TCVitals have a responsibility to work around the inevitable limitations of what forecasters can do for TCVitals in real time. With ensemble systems becoming available, a way to do that will soon be available. However, the TCVitals' limitations must first be quantitatively understood so that model developers can take them into account. That is the motivation for the present study, which compares the TCVitals storm location and intensity to postseason reanalysis values found in the best-track database and statistically compares the TCVitals storm depth to 946 Tropical Rainfall Measuring Mission (TRMM) overpasses. All storms of tropical depression strength or stronger in all basins are analyzed, with a special focus on National Hurricane Center TCVitals for the North Atlantic and eastern Pacific basins, the main areas of responsibility for NCEP. In addition, the sensitivity to TCVitals on the Hurricane Weather Research and Forecasting (HWRF) model is examined by rerunning the 2011 HWRF for the 2010 North Atlantic season twice: once with TCVitals input and once with best-track input.

1. Introduction

The Tropical Cyclone Vitals Database (TCVitals) is an archive of Cyclone Message Files, which contain information such as cyclone location, intensity, horizontal wind and pressure structure, and depth of convection (NCEP 2011; NOAA Developmental Testbed Center 2011), created in real time by forecasting centers. These vitals are used during the vortex relocation and bogusing process in the Global Forecast System (GFS; Liu et al. 2000), Geophysical Fluid Dynamics Laboratory (GFDL; Kurihara et al. 1993, 1998; Bender et al. 2007), and Hurricane Weather Research and Forecasting (HWRF) (NOAA Developmental Testbed Center 2010, 2011)

models. The resulting fields then make up the first-guess fields for data assimilation. This is done both for real-time forecasts and postseason tests of new model versions.

Two other databases, featuring the best tracks from the National Hurricane Center (NHC) and Joint Typhoon Warning Center (JTWC), are used by the National Centers for Environmental Prediction (NCEP) to validate model performance. The best tracks are created and maintained by the same organizations that produce the cyclone message files. The best-track databases contain many of the same storm vitals as the TCVitals database. However, the best tracks are updated postseason (Chu et al. 2002; Avila 2002), while TCVitals is not. TCVitals is a real-time assessment of the cyclone state, which is necessarily based on limited information with strict time constraints. There are therefore differences between the TCVitals and best-track storm location and

Corresponding author address: Sam Trahan, NCEP/EMC, 5200 Auth Rd., Camp Springs, MD 20746.
E-mail: samtrahan@samtrahan.com

intensity that are due to the more complete set of data used by the best-track analysis. The question addressed here is whether the real-time assessment of the cyclone state is accurate when compared to a more complete later analysis for which much more information and time are available.

One specific problem addressed in this paper is to clarify the meaning of the TCVitals storm depth indicator, which is used by the HWRF model in a number of phases of its initialization. Such usage is being phased out of the HWRF due to the storm depth indicator's perceived unreliability, but has not been completely removed from the 2011 operational version of HWRF (NOAA Developmental Testbed Center 2011). In TCVitals, storms are assigned a "D" for "deep convection," an "M" for "medium-depth convection," and an "S" for "shallow convection." This is then used twice in HWRF. First, it is used to specify the depth of the storm size correction when generating the "first guess" as input to data assimilation. Then, the three-dimensional variational data assimilation (3DVAR) version of the gridpoint statistical interpolation (GSI) program (Kleist et al. 2009; NOAA Developmental Testbed Center 2011) is run only for D storms, since preliminary tests have shown that it worsens S- or M-depth storm forecasts. These two TCVitals-based HWRF changes improve model forecasting skill, so this suggests there is some systematic difference between the three depth classifications.

Section 2 briefly examines the technical difficulties brought on by errors in the archiving system used for message files. In section 3 we perform a statistical analysis of the differences between TCVitals and best-track location and intensity results and its impact on HWRF forecasting skill, and section 4 uses hundreds of Tropical Rainfall Measuring Mission (TRMM) satellite overpasses of storms from 1998 through 2010 to analyze the relationship between TCVitals storm depth indicators and objective satellite measurements. Section 5 contains a summary and conclusions.

2. Databases used and quality control processes

a. TCVitals database and cyclone message files

The TCVitals database consists of vitals information for all storms and cyclogenesis cases in 6-h intervals. A cyclone message file (message) consists of vitals information for a single cyclone or cyclogenesis case. All messages sent are intended to be archived in the TCVitals database. In most cases, there are multiple messages for the same storm and time, as the forecaster has multiple opportunities to update these "best guess" vitals. In real

time, the raw message files are used to initialize models, while in postseason retrospective runs, the TCVitals database is used since the message files are no longer available. Unfortunately, the TCVitals files are fraught with numerous data quality issues, especially for earlier years. These issues include nonphysical values, obvious typos, misnamed storms, incorrect analysis times, and otherwise corrupted entries. In addition, some analysis times had vitals messages posted, but the vitals were never archived to the TCVitals database. This is offset by the fact that multiple vitals entries are available at any given time—usually at least one is usable.

Some manual and automatic quality control was performed for this study. The Environmental Modeling Center's (EMC) HWRF group has a copy of the TCVitals where missing vitals were added postseason for the North Atlantic, eastern Pacific, and central Pacific basins. That database was used as the starting point for this analysis. Second, various automatic quality control procedures were applied to that version, changing to a consistent database format, removing entries with nonphysical values, and removing clearly corrupted entries. Third, some manual quality control was performed, partly to verify the automatic quality control, and also to remove additional erroneous lines that were not caught by the automatic process.

Despite the limitations encountered, vitals information for most of the storms from 1997 to 2010 survives the quality control process. This is in part because redundant TCVitals entries are created for the same storm and time, sometimes containing corrections to earlier values. Hence, a statistical analysis of the TCVitals database is still possible. It is also worth noting that, while the recent years of TCVitals still have problems, the problems are comparatively rare, and appear to consist entirely of missing vitals messages.

b. Best-track databases

For this analysis, we use the postseason reanalysis best-track databases from the National Oceanic and Atmospheric Administration/National Hurricane Center (NOAA/NHC) and U.S. Navy Joint Typhoon Warning Center (JTWC). The best-track database used is not the Atlantic Basin Hurricane Dataset (HURDAT) public best-track reanalysis familiar to many researchers. Instead, the original B deck files are used, which contain all cyclones of tropical depression strength and stronger for all basins, and also contain additional fields. This is the format distributed publicly by JTWC (Chu et al. 2002), and is also used internally throughout NOAA.

The best-track data used in this study come from several sources. Analysis of North Atlantic and eastern

Pacific tropical cyclone tracks are the responsibility of NHC. Meanwhile, the Southern Hemisphere, Indian Ocean, and western Pacific are the responsibility of JTWC. Responsibility for the central Pacific is shared between the two centers, so data were taken from NHC when possible and JTWC otherwise. All pre-2010 season JTWC tracks were public JTWC reanalysis tracks (Chu et al. 2002) taken from the JTWC website while 2010 JTWC tracks and all NHC tracks were obtained from internal NCEP sources.

Only numbered storms (tropical cyclones of depression strength or stronger) were used in this study. This was done partly because the pseudonumbers 90–99, which are used for cyclogenesis cases, are reused for multiple different storms in the same season, greatly complicating any analysis. Even if a manual analysis was to be done to separate out different unnumbered storms, the unnumbered best-track data are generally not archived, so such an analysis is not even possible. This has the unfortunate side effect of eliminating many NHC S TCVitals storm-depth vitals cases, since the S depth was used liberally in those cases.

3. Comparison of TCVitals and best-track location and intensity

The NCEP models are initialized with the TCVitals and validated against the best track, so any differences between the two datasets can cause errors or biases in the model initial state. In this section, we compare the TCVitals and best-track values for the storm location, maximum wind, and minimum central sea level pressure per season and basin.

This should not be thought of as a rigorous validation of the TCVitals. A rigorous validation would require comparison to the true storm track and intensity, quantities that are not known. The best information we have is the postseason best-track reanalysis, which is used by NCEP and NHC as the “ground truth” values of storm track and intensity for model validation. Hence, applying the same validation process on the TCVitals makes sense. However, small differences between the TCVitals and best track data be an indication that fewer revisions were done postseason, an important point to consider if one uses these results for improving forecasting.

We analyze the differences between the two databases, using the same statistical measures used when performing postseason analysis of model forecasting skill. In keeping with NHC and NCEP model validation methods, we will attempt a per-season, per-basin analysis when possible. The NHC and NCEP use such a discretized analysis since the synoptic scale changes greatly from

year to year and basin to basin. For some statistics, a combination of multiple years or basins is required in order to have enough samples for statistical convergence. However, throughout this study, NHC and JTWC data will always remain separated.

In this study we define TCVitals “bias” and “error” according to common usage: bias is the average of differences relative to the best track, and error is the average of the absolute value of differences relative to the best track. Similarly, the storm center difference is the average of the great arc distance of the TCVitals storm center location from the best-track location.

Since a postseason analysis is required for this comparison, the only basins we considered for the 2009 season were the North Atlantic and eastern Pacific, for which NHC’s 2010 reanalysis was completed. Note that JTWC defines the Southern Hemisphere basin season in such a way that it actually starts in the middle of the previous year and ends in the middle of the current year. Hence, some storms that actually started in 2009 are classified by JTWC as 2010 Southern Hemisphere season storms and, thus, were discarded to ensure sufficient time for a postseason reanalysis of the entire season.

Two basins are relatively inactive in most years: the Indian Ocean and central Pacific basins. The central Pacific basin will not be discussed in this section since there are too few samples for a per-season analysis. For the sake of completeness, data for that basin will be included; however, definitive conclusions about the central Pacific basin cannot be made at this time. The Indian Ocean will be analyzed, but only results consistent among multiple years will be discussed due to the relative inactivity of that basin.

a. Results of comparison

In Figs. 1a and 1b, we plot the mean difference between the TCVitals and best-track winds, averaged over the entire basin season. Error bars are plotted at one standard error. It is evident that the magnitude of the bias is quite small on the scale of hurricane winds, with biases of negative 0.5–2 kt (meaning a weaker TCVitals storm) for most basins and years ($1 \text{ kt} = 0.514 \text{ m s}^{-1}$). For the most part, biases are negative, with the exception of the Indian Ocean basin, where the bias is typically plus 0–2 kt after 1999, except for 2007.

Figures 2a and 2b show the pressure bias with error bars at one standard error. Biases are around positive 0–0.5 mbar (meaning a weaker TCVitals storm) in most years and basins, with occasional negative biases in JTWC TCVitals and also in the NHC’s eastern Pacific in 2007. The NHC North Atlantic TCVitals data have a consistent positive bias of around 0.3–0.5 mbar from 2003 to 2009.

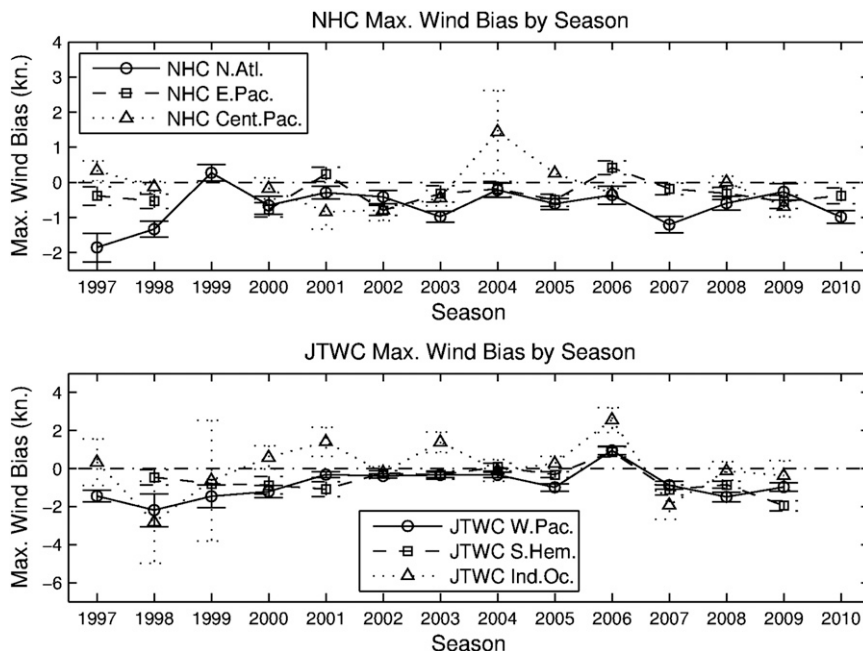


FIG. 1. Per-season, per-basin bias (mean difference) in maximum 1-min-sustained, 10-m-altitude wind estimates: TCVitals relative to the best track: (a) NHC and (b) JTWC.

While 0.5–2 kt of error in a single forecast is small compared to the total storm strength, seasonal biases of ones of knots are considered significant in internal NCEP hurricane model vetting for busy seasons like the North Atlantic during 2008 and 2010. The fact that TCVitals could contribute up to a 2-kt negative bias in

the initial state must be taken into account in future model assessments, in order to isolate errors in the initial state from model errors.

Figures 3 and 4 show yearly north–south and east–west biases. For most basins, the storm center track bias rarely exceeds 5 km. Again, a notable exception is the

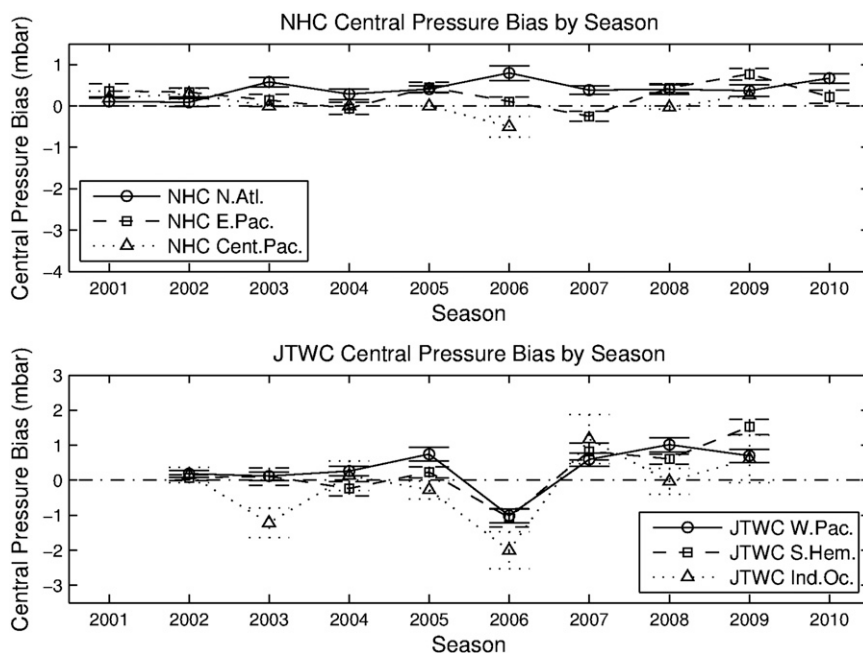


FIG. 2. As in Fig. 1, but for minimum central sea level pressure.

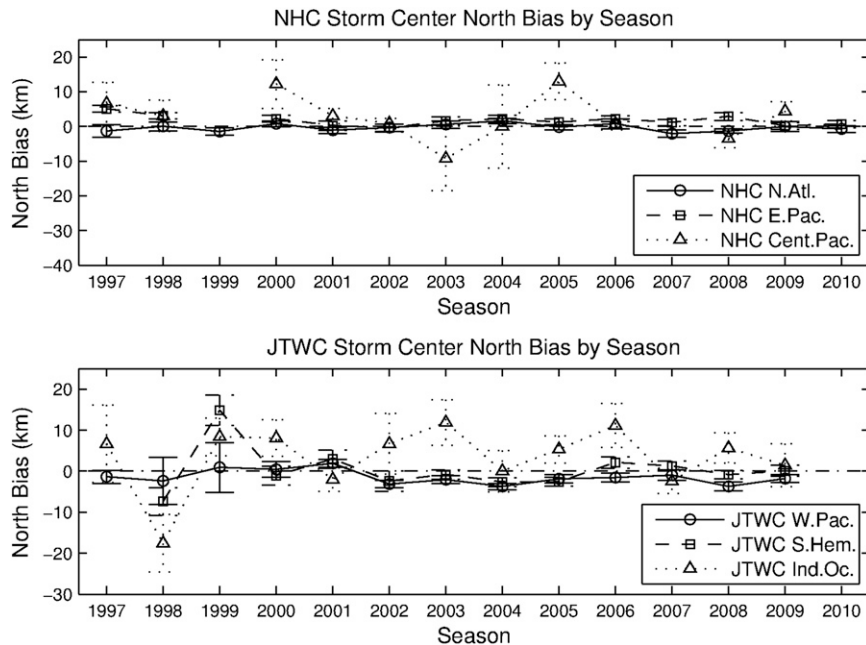


FIG. 3. As in Fig. 1, but for storm-center latitude and longitude (km): (a) NHC north bias and (b) JTWC north bias.

JTWC Indian Ocean TC Vitals, which does display frequent north, west, or northwest biases, even in recent years. Figures 5a and 5b examine cases with large TC Vitals versus best-track differences, and we see that a significant number of such cases are present in all of the relatively active basins. The JTWC vitals appear to

have more large track differences than the NHC vitals, a serious concern because of the well-known possibility of nonlinear amplification of initial errors in a later forecast. Large intensity errors (not shown), greater than 15 kt or 10 mbar, are rare, occurring around 1% of the time in NHC TC Vitals from 1997 to 2010 and 2% of the

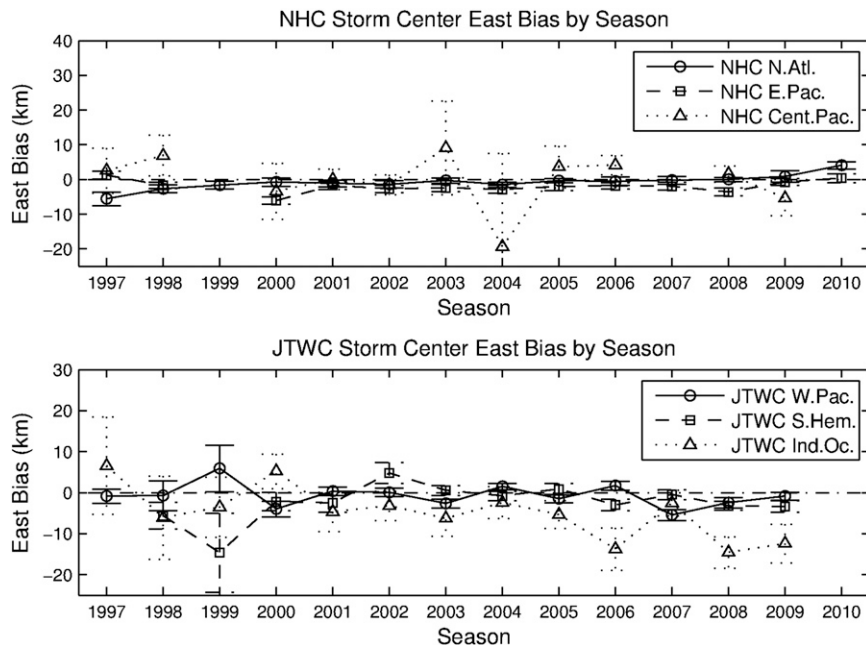


FIG. 4. As in Fig. 3, but for (a) NHC east bias and (b) JTWC east bias.

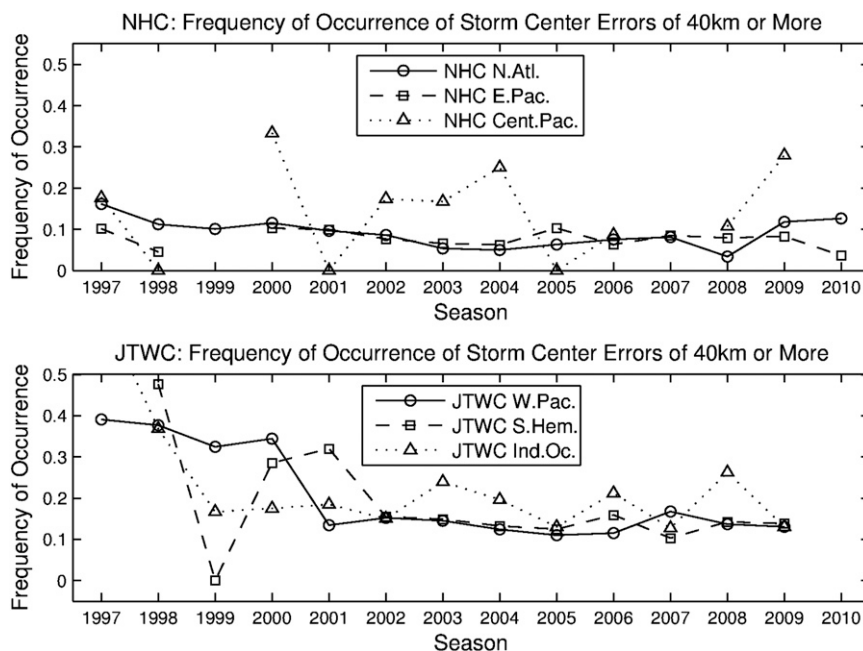


FIG. 5. Frequency of occurrence of large differences between TCVitals and best-track storm-center locations: (top) NHC and (bottom) JTWC.

time in JTWC TCVitals after 2001. This can potentially have a large impact on HWRf simulations due to several cutoff values in their various parameterizations and initialization programs, where a pattern of behavior changes above a certain wind speed or below a certain pressure. However, since the occurrence is so rare, it is expected to have a much lower impact than large track errors.

In all of the plots in this section, it can be seen that the vitals for the two primary basins of interest, North Atlantic and eastern Pacific, are consistently of a higher quality than the other basins. The North Atlantic and eastern Pacific basins do have a consistent negative wind bias of 0.5–1.5 kt; these errors in the initial state could contribute to model forecast errors but are, of course, distinct from model bias. These basins also show large track errors in a few cases, but they occur less frequently than in other basins. The JTWC TCVitals improve drastically after 2001, but the Indian Ocean tracks still have a consistent bias and usually larger errors than the other JTWC basins.

The TCVitals versus best-track storm center differences are larger in weaker storms. This can be seen in Figs. 6a (NHC) and 6b (JTWC), which show scatterplots of the best-track maximum wind against storm center distance. For each of several maximum wind values, lines are drawn at one and two root-mean-square (RMS) storm center distances to show the spread of the distribution. The spread of the distribution is clearly larger for the weaker storms.

b. Impact on forecasting skill

To measure the impact of the TCVitals–best-track differences on forecasting skill, we ran the NCEP 2011 operational HWRf configuration on the NOAA Jet supercomputer for all numbered storms in the 2010 North Atlantic hurricane season twice. The control run is simply the operational configuration, run on the Jet supercomputer. The other is identical except that best-track data were used instead of those from TCVitals. The model is described in detail in the HWRf science document (NOAA Developmental Testbed Center 2011). In total, 389 HWRf simulations from each configuration are used in this analysis (hence 778 total), allowing for a bulk statistical analysis of the season.

In this analysis, we will compare the two HWRf configurations' forecast tracks and intensities out to 120 h to the best-known track and intensity: the NHC best-track values. The definitions are the same as before: bias is the average difference from the best track, and error is the average absolute difference. In all cases, only homogeneous statistics are used; if one model configuration does not reach a given forecast hour for a given storm and forecast cycle, then neither configuration will be considered. In addition, a forecast can only be considered for validation in hours where the best-track storm position and intensity are available. The result of these two restrictions is that fewer forecasts are available at later hours, as can be seen in Table 1.

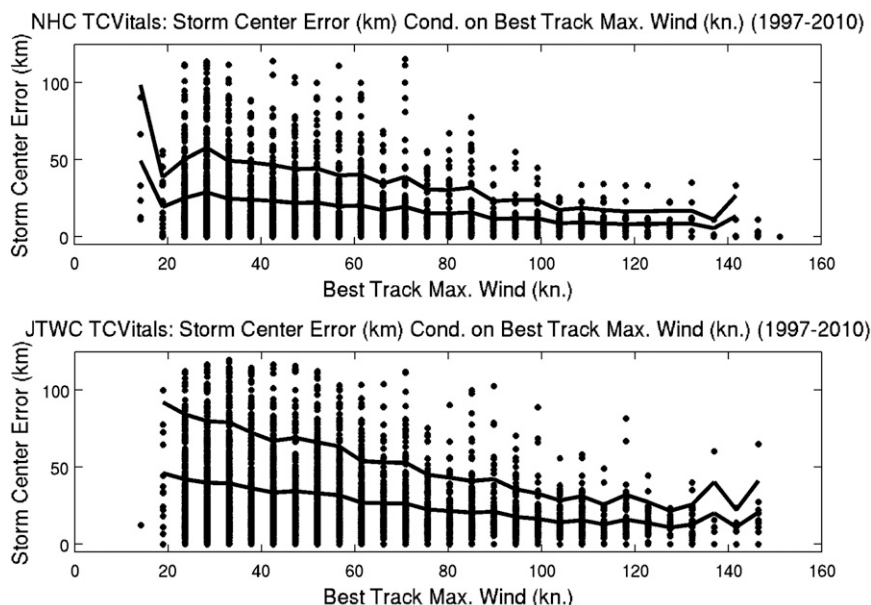


FIG. 6. Scatterplots of the distance of TC Vitals storm center from best-track storm center, against best-track maximum wind. Lines are drawn at one and two RMS error to show the distribution spread. Shown are (a) NHC and (b) JTWC distances.

As examined in the prior section, the NHC 2010 North Atlantic TC Vitals data have a consistent negative intensity bias, and that holds true in the 2010 season as well, with a negative intensity bias of 1 kt (see Fig. 1). Figure 7d shows that the HWRF run with best-track initialization has a bias that is about 0.5–1 kt higher than the TC Vitals-initialized version for most forecast hours, all the way out to the 96-h forecast. The differences in forecast error, on the other hand, are small and vary greatly with forecast hour, as shown in Fig. 7a.

What is perhaps of more interest is the track error statistics. The mean track error (shown in Fig. 7b) shows a small, almost nonexistent, variation between best-track and TC Vitals initializations. The two contributions to that error are the bias and the spread of the distribution, which both have significant differences between the model configurations. The spread of the distribution is measured by the standard deviation in Fig. 7c, which we see to be significantly lower in the best-track-initialized HWRF results than in those of TC Vitals-initialized HWRF. The best-track-initialized HWRF has larger north (Fig. 7e) and east (Fig. 7f) biases at later forecast hours, despite having zero bias at the analysis time.

4. Comparison of TC Vitals storm depth to TRMM measurements

The TC Vitals database contains storm circulation depth indicators for nearly all vitals entries. The indicators

used are S, M, or D for shallow-, medium-, or deep-storm circulation, respectively. NHC has informed EMC developers in the past that these indicators are subjective, may vary from forecaster to forecaster, and are intended only for guidance, not as an objective measure of storm structure (V. Tallapragada 2011, personal communication). This raises the question of whether the storm-depth indicators have any consistent meaning at all.

The HWRF model has some success using these classifiers for tuning, and also to configure the GSI 3DVAR data assimilation (NOAA Developmental Testbed Center 2011). In older versions of HWRF, both of those options result in improvements in forecasting skill, according to preliminary results. The TC Vitals storm depth therefore does provide useful information, and the goal in this section is to examine whether the meaning of NHC's storm-depth indicators can be quantified using estimates of storm depth from TRMM observations.

To evaluate the storm-depth indicator, we use 946 TRMM overpasses of tropical cyclones from 1998 to 2010,

TABLE 1. Number of HWRF cycles from each configuration validated against the best track at each forecast hour using a homogenous sample set.

Forecast hour	0	12	24	36
No. of cases	389	353	302	269
Forecast hour	48	72	96	120
No. of cases	237	184	142	109

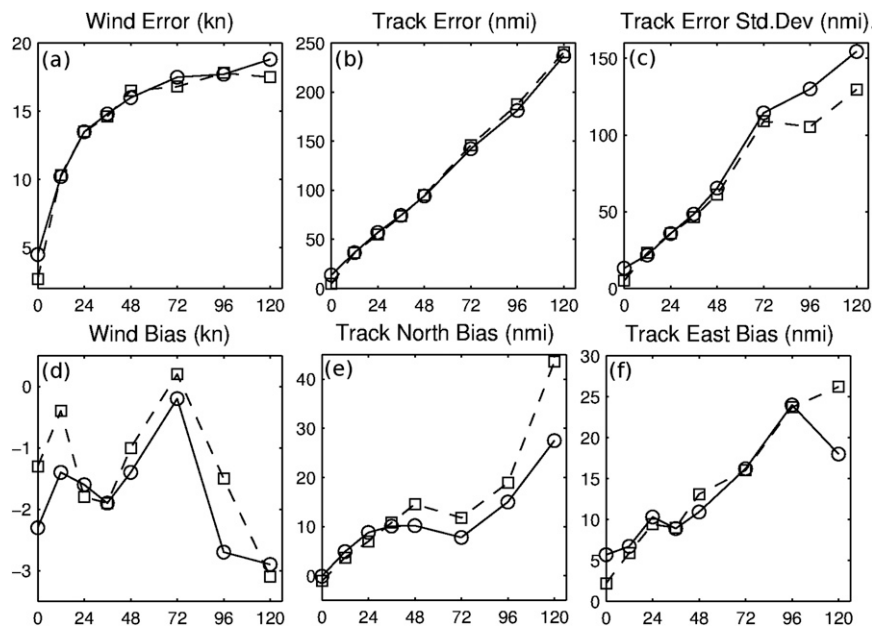


FIG. 7. Bulk statistical analysis of 389 HWRF simulations with TCVitals initialization (solid lines with circles) and 389 simulations with best-track initialization (dashed lines with squares). All plots display bias or error relative to the best track at forecast hours (x axis) 0–120.

selected from times when the satellite ground point passed within 90 km of the best-track storm center at a time when TCVitals data were available. Observations from the TRMM Precipitation Radar (PR) and Visible and Infrared Scanner (VIRS) $10.8\text{-}\mu\text{m}$ channel were then compared to the TCVitals storm depth. TRMM PR provides a measurement of the instantaneous height of convection, while VIRS provides an estimate of the cloud-top temperature of the highest optically thick cloud, even in the absence of active convection. Note that these satellite data provide us with an objective measure of the temperature of the highest cloud, while the TCVitals data provide a subjective measure of the depth of storm circulation. Circulation depth and storm-scale cloud temperature are closely related concepts, since air cannot circulate at high altitudes unless it can reach those altitudes in the first place, and it will not continue to circulate at a high altitude unless it stays there. Hence, this still provides us with a way of objectively analyzing the TCVitals storm-circulation-depth identifier (hereafter referred to as storm depth).

Analysis of JTWC storm depth is restricted to seasons from 2002 onward since the JTWC storm center location accuracy is very low in years prior to 2002. Including years before 2002 in the JTWC analysis worsens the S, M, and D statistics to such an extent that their TRMM cloud-top measurements are nearly indistinguishable. As JTWC's storm center locations have improved drastically

since then, including pre-2002 storm depth in the statistics would not be a fair or meaningful comparison.

The remainder of this section is divided into two parts: the first analyzes three overpasses in detail while the second contains a statistical analysis of all overpasses.

a. Methodology: TRMM compositing technique

This study focuses on analyzing areas within twice the radius from the storm center of the maximum wind (2RMW) in both the TRMM VIRS and PR data. In some cases, the 2RMW region analyzed sometimes lies partly outside of the TRMM PR swath, which has only a 250-km cross-swath width during part of the TRMM lifetime and reduces to a 224-km width after its 2003 orbit boost. A compositing technique is used to avoid overly weighting the inner regions of the storm, which will have more measurements in proportion to their area due to the small swath width.

For our compositing, we divide the storm up into storm-centered concentric circular regions of 40-km radius and use the samples within that region to represent the variability of the entire circle or annulus, including regions that lie outside the swath. The 40-km radius is used because it is approximately three TRMM grid points in the diagonal direction during the years when the TRMM resolution is 5 km, allowing for sufficient TRMM PR sampling of each region, including the inner 40-km radius circle. Larger-radius bins risk mixing statistically

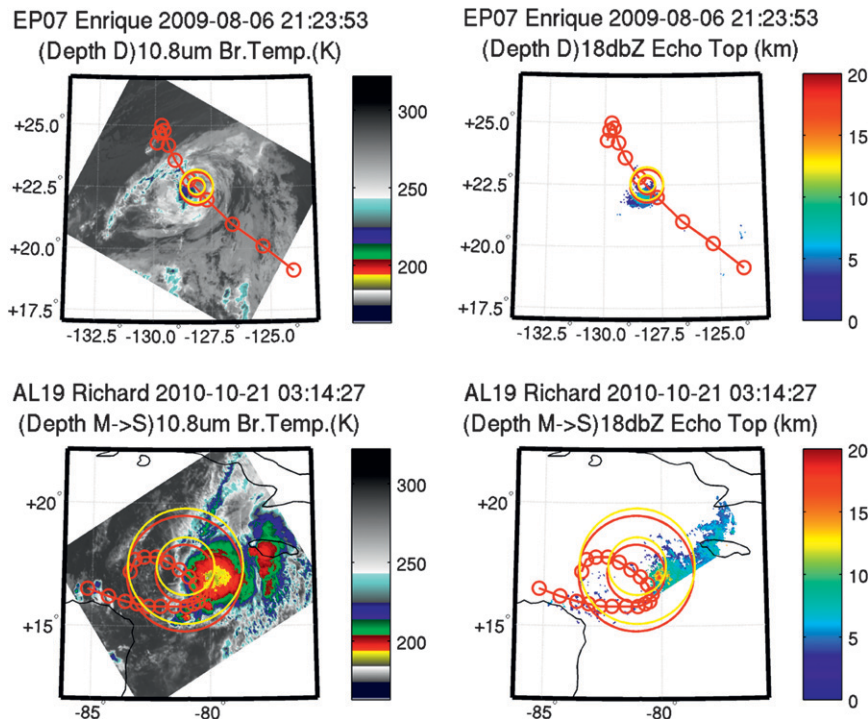


FIG. 8. (top) Tropical Storm Enrique (2009) VIRS 10.8- μm IR channel brightness temperatures and PR storm-top height at 2123:53 UTC 6 Aug 2009. (bottom) Hurricane Richard (2010) VIRS 10.8- μm IR channel brightness temperatures and PR storm-top height at 0314:27 UTC 21 Oct 2010. Best-track positions 48 h before and after crossover are drawn as a sequence of red circles. Two concentric yellow circles are drawn at 1 and 2 RMW around the TCVitals storm-center location, while two concentric red circles are drawn at best-track position at 1 and 2 TCVitals RMW.

distinct storm features into the same region (i.e., outer rainbands and eyewall), while smaller regions result in discarding many storms due to undersampling.

If all regions out to $2 \times$ the radius of maximum winds (2RMW) are well sampled, then we perform the following area-weighted sum to determine the fraction of the total storm area (or total area with radar return):

$$F_{\text{storm}} = \frac{\sum_i \pi (40 \text{ km})^2 [i^2 - (i-1)^2] F_i}{\sum_i \pi (40 \text{ km})^2 [i^2 - (i-1)^2]}, \quad (1)$$

where F_i is the fraction of region i that meets the given criterion and F_{storm} is the fraction of the storm that meets the given criterion.

There is a possible sampling bias that can occur due to this compositing method. The TRMM satellite's inclination limits its PR sampling to within 35°N and 35°S . When the TRMM satellite is viewing close to the 35°N or 35°S region, the satellite track is close to a constant-latitude parallel. Hence, the regions outside of the PR

swath will always be outside to the north or south when the satellite is at the northern or southern parts of its track. At other latitudes, the satellite will pass at two different angles depending on whether it is at the ascending or descending portion of its orbit. However, the farther the satellite is from the equator, the closer those angles will be to one another. These biases should not impact VIRS results due to very few storms having their 2RMW circles lie even partly outside of the VIRS swath.

Before moving on, it is worth examining one common problem in the TCVitals storm-depth indicator. The storm-depth indicator is advertised as the "depth of storm circulation." There are frequent cases of storms with scattered deep convection, but no storm-scale cyclonic circulation. If a storm has deep convection, but no storm-scale circulation, then what is the depth of its circulation: S, M, or D?

Eastern Pacific Tropical Storm Enrique (2009) and North Atlantic Hurricane Richard (2010) illustrate this problem. VIRS and PR data for those storms are shown in Fig. 8. Enrique has convection reaching around 210 K and 8 km. However, that is present only in one tower to

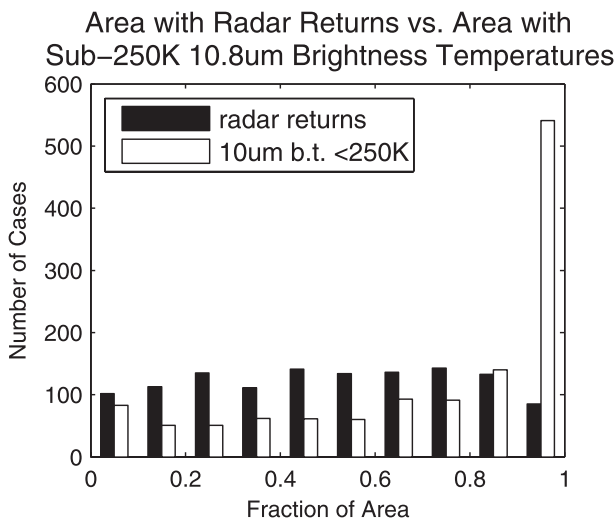


FIG. 9. Using all TRMM overpasses in this study, PDFs of the fraction of the storm area with active radar returns and the fraction of the area with brightness temperatures <250 K.

the south of the storm. The bulk of the outflow is at much lower altitudes with much of the storm having no apparent convection at all. Enrique was classified as a D-depth storm at the TCVitals times immediately before and after this TRMM crossover. At the other extreme is North Atlantic Hurricane Richard (2010), which has an area of likely penetrative convection visible in both the radar and VIRS, but like Enrique, the convection is only in one small region of the storm. However, Richard was variously categorized as M or S during this time.

This suggests that different forecasters have different ways of solving the difficult problem of assigning a depth of storm circulation to a deeply convective storm with no apparent storm-scale circulation. Generally, in the 2008, 2009, and 2010 NHC TCVitals dataset, storms of that sort were given M or S depths, with D less frequent. Because tropical cyclones often spend much of their lifetimes underdeveloped, these sorts of difficult cases represent a significant portion of the whole dataset.

b. Results: Statistical comparison of TRMM and TCVitals

For the TRMM VIRS comparison, we use the N th percentile brightness temperature of the $10.8\text{-}\mu\text{m}$ channel within 2RMW for varying values of N . The storm center and radius of maximum wind used were both taken from the TCVitals dataset to ensure that the same region that the forecaster was analyzing is compared to TRMM. The $10.8\text{-}\mu\text{m}$ channel is chosen because it has a fairly small amount of water vapor absorption, and the brightness temperatures for deep clouds vary only by a few degrees from day to night.

TABLE 2. Number of TRMM PR and TRMM VIRS overpasses for each TCVitals storm depth set.

TCVitals		PR		VIRS	
Center	Depth	Passes	Storms	Passes	Storms
NHC	D	405	203	389	197
NHC	M	154	105	150	108
NHC	S	26	24	34	30
JTWC	D	199	101	198	100
JTWC	M	99	72	96	70
JTWC	S	54	44	55	45

For the TRMM PR comparison, the TRMM 2A23 storm-top height product, based on the 18-dBZ echo-top height, is used. The 18-dBZ height is chosen partly because it is the lowest reflectivity measured by TRMM PR, and also because, above the melting level, 18 dBZ is typically indicative of active convection, especially with reflectivity above 10 km. The statistic used is also different: the N th percentile echo top among the area of the storm that has radar signatures present. This is done because, even in storms with large eyewalls and major rainbands, the region with active convection at any instant makes up a much smaller fraction of the area than the area with clouds.

This effect can be seen in Fig. 9, which contains probability density functions (PDFs) of the fraction of the storm area with active radar returns and the fraction of the area with brightness temperatures below 250 K. In a vast majority of cases, the entire storm has sub-250-K brightness temperatures, while the region with radar returns varies roughly equally from 0% to 100%.

The numbers of PR and VIRS overpasses for the various cases are shown in Table 2. The storms column indicates the number of unique tropical cyclones viewed with that storm depth, center, and device when the device was functioning. The passes column reveals the total number of TRMM overpasses of that storm depth and center when the device was functioning. The number of PR and VIRS vary partly because of varying instrument failures, and partly because some PR overpasses had to be discarded due to too few atmospheric columns with a reflectivity signal, especially in S-depth cases.

We will analyze the relationship between TCVitals storm depth and the VIRS and PR data by looking at PDFs of the 25th percentile VIRS $10.8\text{-}\mu\text{m}$ brightness temperature and 75th percentile PR storm-top height conditioned on TCVitals storm depth. PR data will be binned into 1-km echo-top-height bins, and VIRS data will be binned into 10-K brightness temperature bins. The choices of 25th percentile height and 75th percentile brightness temperature mean that 25% of the innermost

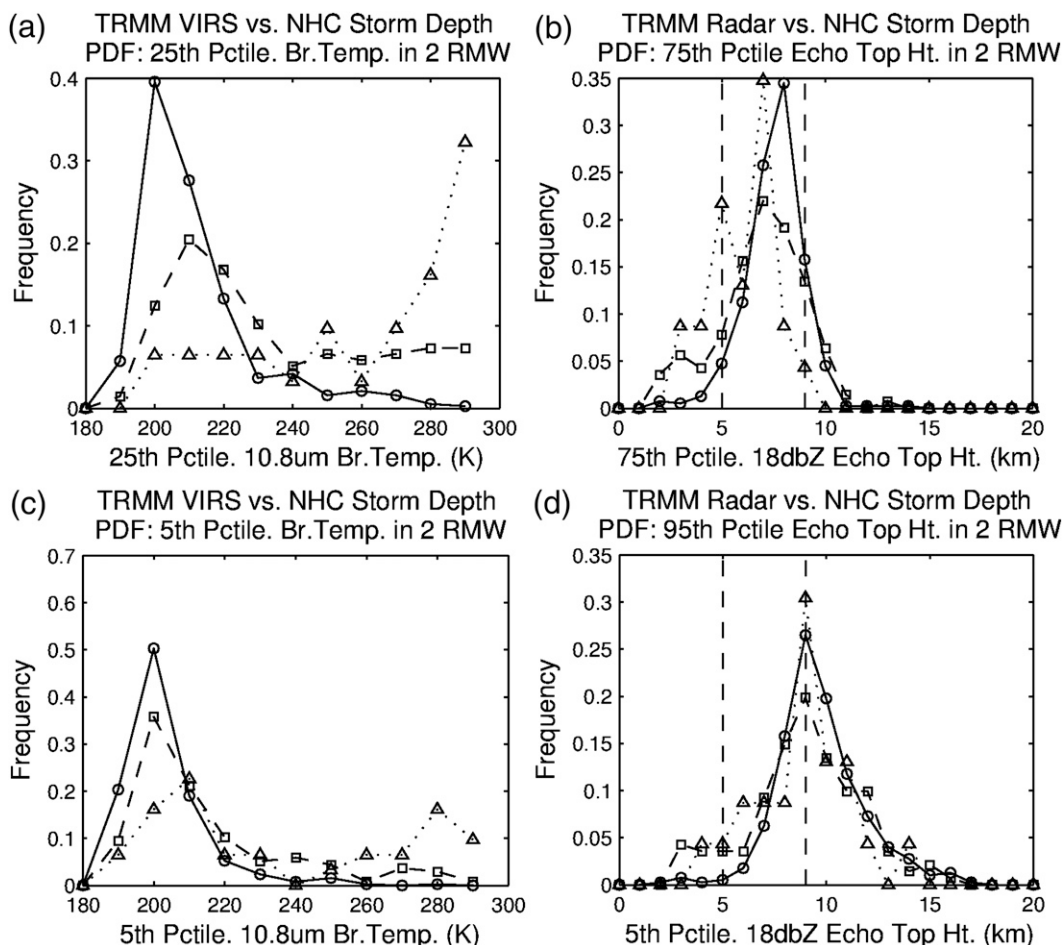


FIG. 10. Histograms calculated from data within 2RMW in TRMM overpasses of NHC storms. Histograms of (a) 25th-percentile VIRS 10.8- μm brightness temperature, (b) 75th-percentile PR 18-dBZ echo-top height, (c) 5th-percentile VIRS 10.8- μm brightness temperature, and (d) 95th-percentile PR 18-dBZ echo-top height. In all plots, solid lines with circles are D depth, dashed lines with squares are M depth, and dotted lines with triangles are S depth.

2RMW region of the storm (hundreds or even thousands of km^2) is at altitudes above that height or temperature.

Figure 10b shows the fraction of storms with a 75th percentile 18-dBZ echo top in each 1-km-height bin, where echo tops have been restricted to those within 2RMW. (Vertical lines are drawn as a guide to the eye: at 5 km, the nominal freezing level, and at 9 km.) Figure 10a depicts the fraction of storms with their 25th percentile VIRS 10.8- μm brightness temperature, which is also restricted to within 2RMW. Results from different TC/Vitals storm-depth classifications are highlighted. This shows a clear difference between the S and D classifications, with the M classification containing some elements of both S and D. No storms classified as S had 75th percentile echo tops greater than 9 km, and few D storms had 25th percentile brightness temperatures

higher than 250 K or 75th percentile echo tops lower than 5 km, the nominal freezing level.

This is explainable by a possible sampling bias in the data available to forecasters in real time: infrared data are often the only high-resolution cloud-top measurement available for storms, so it is not surprising that VIRS results compare well to the TC/Vitals storm depth. However, infrared can only see to the top of the highest cloud. If the highest cloud is due to convection ending an hour or more earlier, then the bulk of the current convection may be much lower. TRMM PR, being a Ku-band radar, is able to see through the upper outflow to the instantaneous height of the convective cells below. Hence, sudden changes in the instantaneous convection height may not yet show up in any infrared data forecasters have available, while it will show up in radar. This problem is mitigated somewhat by the fact that

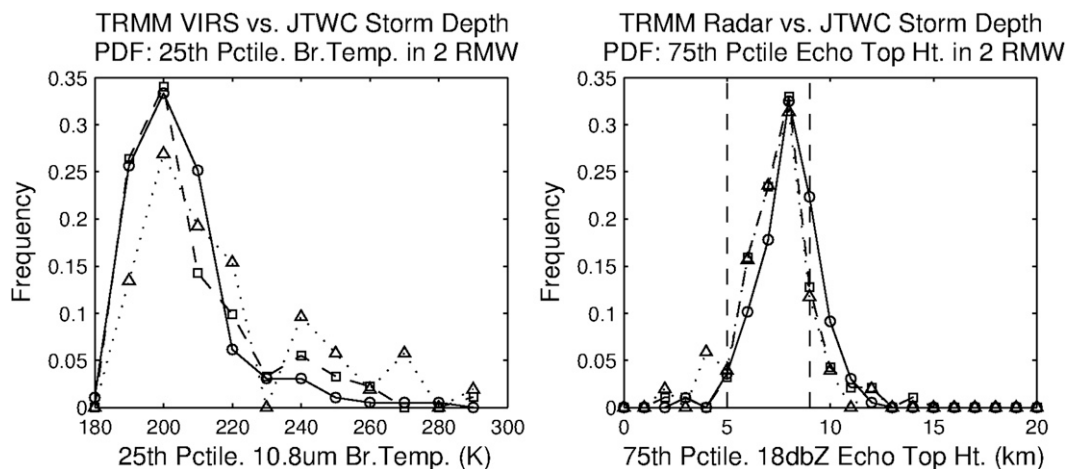


FIG. 11. As in Fig. 10(a),(b), but for JTWC storms.

passive microwave imager measurements are available about twice a day.

When examining the JTWC TCVitals data, the above-mentioned SMD patterns break down, as shown in Figs. 11a and 11b. There is no clear difference between the S, M, or D datasets' VIRS results; the S-, M-, and D-conditioned PDFs are nearly identical in brightness temperature, other than a very slight tendency of S storms toward higher brightness temperatures. In the PR data, M and D storms rarely have sub-5-km 25th percentile echo tops. However, the PDFs of S, M, and D are otherwise very similar, aside from a slight tendency of D storms toward higher-altitude echo tops.

The 25th percentile brightness temperature (75th percentile echo-top height) was chosen for these figures since it is the lowest percentile brightness temperature (highest percentile echo top) that still produces the same patterns discussed above. When the brightness temperature percentile (echo-top percentile) was varied toward the median, the NHC data do not exhibit any significant changes in the patterns of S, M, or D: S storms still rarely have high N th percentile echo tops or low brightness temperatures, D events rarely have low N th percentile echo tops or brightness temperatures, and M storms remain a “we are not sure” category.

However, if one varies toward extreme convection, the NHC PDFs begin to line up with one another and the S–M–D distinction is lost. This can be seen in Figs. 10d and 10c, which show histograms of the echo-top height and brightness temperature, conditioned on storm-depth classifier. This is an expected result, however, since the TCVitals storm depth is supposed to be the depth of the storm circulation. The height of the circulation is better classified by what a quarter of the storm is doing, rather than by the height of a few isolated towers.

5. Conclusions

In a study of 1997–2009 vitals data, the National Hurricane Center's North Atlantic and eastern Pacific TCVitals location and intensity values are found to be of fairly high quality, with some caveats. A negative bias of 0.5–1.5 kt is found in the maximum wind estimates. Biases of several knots in mesoscale models are considered unacceptably large, so evaluations must take into account that the input state from TCVitals has a small bias as well. A 0.5–1.5-mbar positive pressure bias that corresponds to the negative wind bias is also seen. Track biases for all years are less than 5 km in any direction. Track errors of more than 40 km are found in around 10% of the cases, throughout all years of the database. RMS track errors are seen to be larger for weaker storms.

The Joint Typhoon Warning Center's western Pacific and Southern Hemisphere TCVitals data in years after 2001 have poorer agreement with the best-track results. A track bias relative to best track is seen in most years in the Indian Ocean basin, ranging from 5 to 15 km in each of the west and north directions. Mean track errors relative to best track are slightly higher than in NHC: around 20 km compared to 15 km for NHC. Track errors of more than 40 km are also slightly more frequent: around 15% of the time for both basins. Pressure bias varies widely from year to year, and the maximum wind bias is typically positive at 0–2 kt. JTWC, as with NHC, has larger track errors for weaker storms.

To examine the sensitivity of forecasting skill to TCVitals, two sets of 389 HWRF forecasts were performed for the 2010 North Atlantic season, one set with TCVitals initialization, and one set with best-track initialization. Most forecast hours have a 0.5–1-kt more positive wind bias, including the 120-h forecast, which

is consistent with the 1-kt negative bias of the TCVitals for that basin and season. The mean track error between the two configurations is similar, but the two components of the track error—bias and standard deviation—are not. When running with the best-track initialization, HWRF develops an increase in its north bias at 48 h and later of 5–15 n mi, and a 10-kt-larger east bias at 120 h. To counter that, the track standard deviation decreases at 96 h and later.

As for TCVitals storm depth, the NHC does a good job of eliminating extremely deep storms from the S-depth cases, and extremely shallow storms from D-depth cases. However, the M classification spans the entire range of cases, from likely tropopause penetrative storms all the way down to storms with little or no active convection. Similarly, there is significant overlap between D and S cases in the middle range of both radar and infrared convective top measures. All of this suggests a rule of thumb for NHC Storm depth: D rarely contains shallow storms, S rarely contains deep storms, and M is a “we are not sure” category; medium storms are distributed among all three categories.

In addition, the NHC height classification seems to be well correlated only with the large-scale convection height; when one analyzes the small-scale convection height 5th percentile convection tops, little difference is seen between S, M, and D.

The only pattern is seen in the choice of storm depth in the JTWC TCVitals data is that D and M storms rarely have 25th-percentile 18-dBZ echo-top heights below 5 km. Other than that, there is no significant difference between the JTWC S-, M-, or D-storm depths in either the TRMM PR or VIRS 10.8- μm channel for any of the statistics chosen. More subtle differences in JTWC classifications may be seen from a much larger sample set or different statistical measures, but with the several hundred cases available from TRMM overpasses and statistical measures used, no difference is apparent. This conclusion remains valid even if the discarded pre-2002 overpasses are included, or if central Pacific or Indian Ocean basin storms are removed (though those plots are not shown in this paper for brevity).

Finally, there were severe and frequent data quality issues in earlier years of the TCVitals database, especially for JTWC vitals. These include problems such as misnamed storms, hours with missing message files that were available in real time, incorrectly formatted lines, or nonphysical quantities. In the 2010 TCVitals data, most of these issues have been resolved, but there are still occasional NHC message files that are not recorded to the TCVitals files.

This study has focused on quantifying the differences between the TCVitals and best-track datasets, and has

given some quantitative insight into the meaning of the TCVitals storm-depth classification. However, there is still much work to be done in evaluating the quality of this database. There are many other vitals quantities this paper does not evaluate, including environmental pressure; radius of outermost closed isobar; 34-, 50-, and 64 kt wind radii; and many others. In addition, no attempt has been made to compare the location or intensity to observations. It is hoped that other researchers will begin to contribute to the evaluation of this database, and that operational data assimilation will improve to the point that the TCVitals-based bogusing is no longer needed.

Acknowledgments. We thank Jordan Alpert, Qingfu Liu, Robert Tulaya, and Vijay Tallapragada at EMC for providing information about the EMC HWRF, GFS, and GFDL models. Also, the University of Maryland, Baltimore County (UMBC), provided us with much needed internet-connected, high-performance computing capacity required for performing the TRMM aspects of this work.

REFERENCES

- Avila, L. A., 2002: Best track determination at NHC. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 11C.1. [Available online at <http://ams.confex.com/ams/pdfpaper/36320.htm>.]
- Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane–ocean prediction system and a summary of its performance. *Mon. Wea. Rev.*, **135**, 3965–3989.
- Chu, J.-H., C. R. Sampson, A. S. Levine, and E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945–2000. Naval Research Laboratory Tech. Rep. NRL/MR/7540-02-16, 28 pp.
- Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705.
- Kurihara, Y., M. A. Bender, and R. J. Ross, 1993: An initialization scheme of hurricane models by vortex specification. *Mon. Wea. Rev.*, **121**, 2030–2045.
- , R. E. Tuleya, and M. A. Bender, 1998: The GFDL hurricane prediction system and its performance in the 1995 hurricane season. *Mon. Wea. Rev.*, **126**, 1306–1322.
- Liu, Q., T. Marchok, H.-L. Pan, M. Bender, and S. Lord, 2000: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. NWS Tech. Rep. 472, 7 pp.
- NCEP, cited 2011: Format of the tropical cyclone vital statistics records (“TCVitals”). NCEP/Environmental Modeling Center. [Available online at http://www.emc.ncep.noaa.gov/mmb/data_processing/tcvitals_description.htm.]
- NOAA Developmental Testbed Center, 2010: February 2010 HWRF users guide. Tech. Rep., 88 pp.
- , 2011: February 2011 HWRF scientific documentation. Tech. Rep., 80 pp.