12C.2 A NEW HISTORICAL DATABASE OF TROPICAL CYCLONE POSITION, INTENSITY, AND SIZE PARAMETERS OPTIMIZED FOR WIND RISK MODELING

Jonathan L. Vigh,* Eric Gilleland, and Christopher L. Williams
Research Applications Laboratory, NCAR,† Boulder, Colorado

Daniel R. Chavas
Earth, Atmospheric, Planetary Sciences, Purdue University, West Lafayette, Indiana

Neal M. Dorst
Hurricane Research Division, Atlantic Oceanographic and Marine Laboratory, NOAA, Miami, Florida

James M. Done and Greg J. Holland
Mesoscale & Microscale Meteorology Laboratory, NCAR,† Boulder, Colorado

Barbara G. Brown
Research Applications Laboratory, NCAR,† Boulder, Colorado

1. INTRODUCTION

Striking advances have occurred in the realm of tropical cyclone (TC) wind observations over the past quarter century. One of the early key advances has been the advent of highly accurate GPS-based navigation for aircraft-based observing platforms. More accurate knowledge of the aircraft position has allowed for considerably more accurate determination of the earth-relative wind speeds. GPS navigation has also resulted in improved center fixing. Other important advances in aircraft-based wind observations include airborne Doppler radar, GPS dropsondes, improved wind retrievals from satellite scatterometers, and the Stepped Frequency Microwave Radiometer (SFMR) instrument for measuring surface wind speeds. In the near future, there is potential for additional novel observing platforms and instruments to come online, such as aerial unmanned vehicles and space-borne geostationary Doppler radar. Since wind risk modeling depends crucially on having the best possible knowledge about the characteristics of the surface wind field, substantial benefits can accrue if these older and newer wind observations across the modern era can be combined in a consistent and transparent manner.

*Corresponding author address: Jonathan Vigh, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000; e-mail: jvigh@ucar.edu

1.1 *Characteristics of the new historical database*

The present work aims to do just this by constructing a new historical database called the **Tropical Cyclone Observations-Based Structure Database** (or **TC-OBS Database** for short). The central focus of the TC-OBS Database is to provide wind information that is optimized for parametric wind risk applications. This is accomplished by providing objective, observations-based estimates for the key parameters of track, intensity, and size (radial extent of winds of a given threshold). Additionally, TC-OBS provides estimates of the radius of maximum winds (RMW), a quantity that has not previously been included as a "best-tracked" quantity in existing datasets. TC-OBS also provides time-dependent uncertainty bounds on the estimates of these four main quantities. Compared with existing historical databases such as the Hurricane Database (HURDAT2, see Landsea and Franklin 2013), TC-OBS provides track, intensity, and radius information at higher spatial and temporal precision. Whereas HURDAT2 rounds track points to the nearest tenth of a degree, intensity to the nearest 5-kt increment, and wind radii to 5-nm or even 10-nm increments, TC-OBS does not round any of its estimates to artificial thresholds. While HURDAT generally provides parameters every six hours, TC-OBS provides estimates of all parameters for each hour as well as any of the off-synoptic time points included in HURDAT2. Like HURDAT2, TC-OBS provides estimates of the 34-, 50-, and 64-kt wind radii (size), but TC-OBS also adds estimates for the radial extent of Category 2, 3, 4, and 5 wind speed thresholds (83-kt, 96-kt, 113-kt, and 135-kt). When data coverage are sufficient, TC-OBS also

includes estimates of the azimuthal mean wind speed. Finally, TC-OBS is setup to include alternative metrics beyond the traditional metrics for intensity and size. One such metric is related to the spatial and temporal coherence of the location of wind maxima.

1.2 *Database coverage*

The TC-OBS Database currently includes data for all TCs that occurred in the North Atlantic basin from 1999 to 2013. Somewhat more than half of all Atlantic TCs were sampled by aircraft frequently enough to provide useful input data for TC-OBS. Additionally, since aircraft reconnaissance generally commences when TCs move west of 55 deg in the basin, TC-OBS does not provide observational refinement for TCs in the eastern half of the basin. For the most part however, there is good aircraft coverage during nearly all periods in which TCs were threatening land, so TC-OBS provides observational refinements for nearly all of the impactful landfalls in the basin. The current version of TC-OBS provides observational refinements for 253 of the 416 TCs that occurred during 1999 to 2013.

1.3 *Scope of this extended abstract*

The focus of this extended abstract is to provide an overview of the methods that have been used to construct the TC-OBS Database. Since the recorded conference presentation and supplementary pdf of the presentation file provide a number of graphical comparisons between the TC-OBS Database parameters, the HURDAT2 parameters, and the underlying aircraft-based observations, these plots are not provided here.[1]

2. **PREPARATION OF OBSERVATIONAL DATA**

In order to make objective-based estimates from observational data sources, it is imperative to clean the data as much as possible, yet not to introduce biases during the process. Toward this goal, three new research-grade datasets were constructed by systematically gathering and standardizing all readily available observations and undertaking extensive quality control efforts. These datasets include: (i) the Enhanced Vortex Data Message (VDM+) Dataset (Vigh 2015b), the QuikSCAT Tropical Cyclone Radial Structure (QSCAT-R) Dataset (Chavas and Vigh 2014b), and the Extended Flight Level Dataset for Tropical Cyclones (FLIGHT+; Vigh et al. 2014).

---

[1]The interested reader is invited to access the recorded presentation and supplementary pdf from the conference program at: https://ams.confex.com/ams/32Hurr/webprogram/Paper293910.html.

Each dataset has been made available to the research community to support a wide variety of research uses beyond the present work[2]. Since the scope of the current document does not permit a full description of these datasets, the reader is referred to documentation that has already been published for the QSCAT-R and VDM+ datasets (Chavas and Vigh 2014a; Vigh 2015a). Documentation for the FLIGHT+ Dataset is still in preparation, but the reader can find considerable detail about the data processing and quality control measures in Vigh (2014).

3. **METHODS**

The construction of optimal estimates of TC parameters from relatively sparse observations is a challenging and interesting problem. The following philosophical considerations have shaped the efforts to devise the objective methods used in the TC-OBS Database:

1. *In situ* aircraft observations from the modern observing period are generally of high enough quality that they are treated as the "gold standard". Therefore, when sufficient aircraft data are present, these values should generally be trusted and given a much higher weight than the background value.

2. When many aircraft observations are available in the analysis window for a given time point, the TC-OBS parameter estimate should be a blended average of the available observations, subject to weighting according to whatever criteria is important for the parameter being estimated.

3. When aircraft data are sparse, the TC-OBS parameter should relax back to the background value. The background value is taken to be the HURDAT2 parameter value, which have been interpolated to the same time stencil as the TC-OBS Database. This ensures consistency between TC-OBS and HURDAT2 when there are not sufficient data to provide observational refinement.

4. Although there is always considerable potential for under-sampling (especially for the intensity parameter), it is normally impossible to determine if a given point is in fact under-sampled or whether it may actually be close to the true estimate of the storm. Therefore, the TC-OBS methods currently make no explicit adjustments for under-sampling.

---

[2]Users may download the datasets and associated documentation at http://verif.ral.ucar.edu/tcdata/.

5. All estimates should be provided at full precision. Since an uncertainty estimate is provided separately, there is no need to round values to arbitrary thresholds.

6. While more complex methods could be envisioned (e.g., Bayesian inference or variational-based data assimilation-type approaches), the philosophy has been to keep the methods simple to allow one to readily understand how the value has been obtained. Thus, the simple methods developed in this initial version of the database can serve as a baseline from which to compare the innovations of more complex methods in the future.

3.1 *General algorithmic approach*

In view of the above considerations, the optimal estimation of each of the main database parameters (track, intensity, RMW, and size) are computed using the following general algorithmic steps.

- **Step 1: Filter/merge observations to eliminate conflicting and/or duplicatory information, keeping the best observations.** Given that observations are being brought in from two different aircraft datasets (the VDM+ Dataset and the FLIGHT+ Dataset), it is necessary to merge the data so that double-weighting does not occur. Objective cutoffs were determined for each observation type in order to merge the data into one time series that contains the best or most reliable observations for the given parameter. As an example, for track points, the Chelmow/Willoughby (C/W) wind centers are judged to have high accuracy than the real-time fixes reported in the Vortex Data Messages (VDMs). Thus, if a VDM fix is available within 30 minutes of a C/W wind center fix, the VDM fix is eliminated and only the C/W wind center fix is used to estimate the TC-OBS value. For the other wind parameters, the filter/merging step also involves reducing flight level wind speeds to surface equivalents. In this initial version of TC-OBS, the flight-level-to-surface reduction factors are based on the values reported by Franklin et al. (2003). For the RMW parameter, the radius of flight level winds is reduced to a surface equivalent radius value using the 0.875 factor reported by Powell et al. (2009).

- **Step 2: Gather relevant data for each time point to be estimated by traversing the available observations using a moving analysis window centered on the target time to be estimated.** By doing this, the observations that lie within the analysis window relevant to a given time are obtained. The difference in time between each observation and the target time is then computed. This means that observations further away in time than 6 h will not contribute at all to the optimally-estimated value. The choice of the analysis window half-width time is a necessary compromise between maintaining "sharpness" and ensuring sufficient data are available to obtain a robust estimate. For most of the key parameters, a half-width time of 6 h was found to give good results. For track, the half-width time is taken to be 8 h.

- **Step 3: Determine the effective number of data points by determining a provisional data weight for each observation and then summing these provisional weights for all observations within the analysis window.** Using nearness-in-time and optional additional "goodness" criteria, the *effective* number of observational data points is determined for each target time. An e-folding basis function is used to give the highest provisional data weight to observations that are nearest to the target time point. Thus, points that are near the edge of the analysis window are given considerably provisional lower weights than points closer to the target time. The provisional data weights are computed using the following formula: $w_{\text{provisional data weight}} = \exp(-\frac{\delta t}{\lambda_{\text{observation influence}}})$, where δt is the absolute difference in time between the target time and the observation time, and $\lambda_{\text{observation influence}}$ is the e-folding time scale for observational data influence. For most parameters, TC-OBS uses $\lambda_{\text{observation influence}} = 4$. As an example, if an observation happens to be at the same time as the target time point being estimated, this formula gives it an effective data weight of 1.000. If the target time point is ± 1 h from the target time, the effective data weight is 0.794. For a point 2 h away in time, the weight is 0.607. For a point 4 h away, the weight drops to 0.368. A point at the very edge of the analysis window has a weight of 0.223. For estimating intensity and RMW, additional "goodness" criteria are used to inform the weight of each observation. Since aircraft typically sample the storm following a figure-4 pattern, a series of passes through the storm may alternatively sample the strong and weak sides of an asymmetric storm. Since most storms are asymmetric, this results in a large scatter, but since the goal is to estimate the maximum surface wind anywhere in the storm, it is the upper bound of the observed values that should contribute most to

the estimate. Thus, the "goodness" criteria gives much higher weights to observations that are near the time-trended upper bound of wind speeds. In this way, observations from the weak side of an asymmetric TC are given little weight, while the observations from the strong part of the TC contribute nearly all of the weight. Similarly, the TC-OBS RMW estimate keys off of the radius of the strongest winds and essentially ignores the influence of local wind maxima that do not contribute substantively to the time-trended upper bound of wind speed.

- **Step 4: Compute total observational and background weights, giving higher collective weight to the observations when the number of effective data points is high, and higher weight to the background value when the number of effective data points is low.** Using the effective number of data points, the total weight of the observations is computed using another inverse e-folding basis function: $w_{\text{background}} = \exp(-\frac{n_{\text{effective data points}}}{\lambda_{\text{background influence}}})$,

  where $n_{\text{effective data points}}$ is the number of effective data points, and $\lambda_{\text{background influence}}$ is the e-folding scale for background data influence. The collective observational weight is then taken to be the residual: $w_{\text{combined observations}} = 1 - w_{\text{background}}$. For the initial version of TC-OBS, an e-folding scale for background data influence is set to $\lambda_{\text{background influence}} = 0.666667$. This value results in a steep drop-off in the background weight (and consequently, a higher weight given to the observations) as the number of effective data points increases. When the number of effective data points is 0.0, the background weight is 1.000. When the number of effective data points is 0.4, the background weight is 0.513. For 1.1 effective data points, the background weight is 0.189. For 2.0 effective data points, the background weight is 0.050. For 4.0 effective data points, the background weight is 0.002 (essentially nil).

- **Step 5: Optimally estimate the parameter value as a weighted average of the observations and the background value**. Once the combined observational weight has been determined in the above step, the individual weights for each observation can be determined by normalizing the provisional weight previously computed for each observation in Step 3 by the combined observational weight computed in Step 4. Then the parameter value is optimally estimated as the weighted average of each observation within the analysis window and the background value at the target time.

The above approach, which can be described as an criteria-informed weighted average, is used to optimally estimate the intensity, RMW, and wind radii. For track points, a somewhat different combinatory approach has been used. For track, once the "good" aircraft fixes have been selected/merged, the resulting array of lat/lon points is supplemented with Best Track points whenever any gap of 3 h between observational fixes occurs. In this way, the TC-OBS track will relax smoothly back toward the HURDAT track when fixes are sparse. An additional difference for track is that instead of combining points using a weighted average, an interpolatory cubic spline is used to determine the track points at the TC-OBS time stencil. Using an interpolatory spline essentially treats the fixes as "truth" and forces the spline to pass through them. This is why it is imperative to eliminate any VDM fixes that are near-in-time to the C/W wind center fixes. Failure to do so results in spurious excursions of the interpolating spline.

3.2   *Time-dependent uncertainty bounds*

The approach to computing the time-dependent uncertainty bounds for each parameter is similar in principle to the methods for the optimal estimate of each parameter, except that rather than computing a criteria-informed weighted average, a criteria-informed weighted *variance* is computed, in which the deviations depend on the characteristic uncertainty of each observation according to its type. By way of example, for intensity, the characteristic uncertainty of *flight level wind observations* results mainly from the uncertainty associated in reducing the flight level wind speed to a surface equivalent value. This uncertainty is quite large and tends to dominate other potential sources of uncertainty. TC-OBS uses the standard deviations given by Franklin et al. (2003) to determine the uncertainty for each flight level observation. As a result of using standard deviations, the flight level uncertainty scales with the absolute value of the flight level wind speed. Thus, reduction factors are noticeably higher for more intense TCs than for weak TCs. For SFMR surface wind observations, Uhlhorn et al. (2007) reported a root-mean-square error of $4\,\text{m s}^{-1}$ that did not depend strongly on the wind speed value. Thus, the uncertainty for SFMR surface winds is taken to be $4\,\text{m s}^{-1}$. One impact of this is that TCs sampled by SFMR have a noticeably smaller estimated uncertainty than for TCs without SFMR.

The above description for intensity uncertainty also translates to RMW and wind radii. Since the flight level data are sampled along each radial leg to determine the maximal radial extent of winds of a given threshold, the characteristic uncertainty associated with flight level-to-surface reduction translates into a range of plausible radii

of extent for the given wind speed threshold. Likewise, the $\pm 4\,\mathrm{m\,s^{-1}}$ uncertainty of SFMR data also result in a range of radii for the radial extent of the given wind speed threshold.

4. SUMMARY

This extended abstract has provided an overview of the effort to build TC-OBS, a new historical database optimized for wind risk modeling. The new database features higher temporal and spatial precision by using the high resolution wind centers obtained from the Willoughby-Rahn center finding method. These wind centers characterize the actual wind center of the TC rather than the geometric center. Since the wind radii and maximum wind locations are referenced from the flight level wind center, the TC-OBS Database track points provide a more accurate reference point for wind risk applications. Data for all metrics is provided at 1-h intervals without rounding to arbitrary thresholds. TC-OBS also includes time-dependent uncertainty bounds based on the inherent uncertainty of the available observing platforms/instruments that were available over the characteristic influence period for each time point. The database also includes alternative metrics beyond the traditional metrics of intensity and 34-, 50-, and 64-kt wind radii (size). Such metrics include the azimuthal mean wind speed, the radial extent of Category 2, 3, 4, and 5 wind speed thresholds (83-kt, 96-kt, 113-kt, and 135-kt), and a metric related to the spatial and temporal coherence of the location of wind maxima.

We expect that TC-OBS will have wide utility for wind risk modeling and calibration of catastrophe models, however many other scientific applications can be envisioned, especially for any researchers needing a high quality database of RMW information. In a future publication, we will provide a validation study that examines how the observations-based refinements of TC-OBS impact return periods for TCs, as well as examine several case studies of well-known impactful landfall events.

The TC-OBS Database is currently slated for release to the wider research community in November 2016.

References

Chavas, D. R. and J. Vigh, 2014a: QSCAT-R: The QuikSCAT Tropical Cyclone Radial Structure Dataset. NCAR Tech. Note NCAR/TN-513+STR, 25 pp., Boulder, CO. doi:10.5065/D6J67DZ4, [Available online at http://nldr.library.ucar.edu/collections/technotes/TECH-NOTE-000-000-000-882.pdf].

Chavas, D. R. and J. L. Vigh, 2014b: QSCAT-R: The QuikSCAT Tropical Cyclone Radial Structure Dataset (Version 1.0). Tropical Cyclone Data Project, National Center for Atmospheric Research, Research Applications Laboratory, Boulder, Colorado. [Available online at: http://dx.doi.org/10.5065/D65B00J3] Accessed 16 Dec 2015.

Franklin, J. L., M. L. Black, and K. Valde, 2003: GPS dropwindsone wind profiles in hurricanes and their operational implications. **18**, 32–44.

Landsea, C. W. and J. Franklin, 2013: Atlantic hurricane database uncertainty and presentation in a new database format. **141**, 3576–3592, doi:10.1175/MWR-D-12-00254.1.

Powell, M. D., E. W. Uhlhorn, and J. D. Kepert, 2009: Estimating maximum surface winds from hurricane reconnaissance measurements. **24**, 868–883.

Uhlhorn, E. W., P. G. Black, J. L. Franklin, M. Goodberlet, J. Carswell, and A. S. Goldstein, 2007: Hurricane surface wind measurements from an operational stepped frequency microwave radiometer. **135**, 3070–3085.

Vigh, J., 2015a: VDM+: The Enhanced Vortex Data Message Dataset: Structure, intensity, and environmental parameters from Atlantic tropical cyclones. NCAR Tech. Note NCAR/TN-517+STR, 72 pp., Boulder, CO. doi:10.5065/D6PR7T26, [Available online at http://nldr.library.ucar.edu/collections/technotes/TECH-NOTE-000-000-000-887.pdf].

Vigh, J. L., 2014: Development of an HWRF diagnostics module to diagnose intensity and structure using synthetic flight paths through tropical cyclones. Final report to the Development Testbed Center Visitor Program, 44 pp., Boulder, CO. [Available online at http://www.dtcenter.org/visitors/reports_2012/DTCVP_FY2012_Vigh_SyntheticProfiles_FINAL_REPORT-13Oct.pdf].

Vigh, J. L., 2015b: VDM+: The Enhanced Vortex Data Message Dataset (Version 1.131). Tropical Cyclone Data Project, National Center for Atmospheric Research, Research Applications Laboratory, Boulder, Colorado. [Available online at: http://dx.doi.org/10.5065/D61Z42GH] Accessed 16 Dec 2015.

Vigh, J. L., et al., 2014: FLIGHT+: The Extended Flight Level Dataset for Tropical Cyclones (Version 1.0). Tropical Cyclone Data Project, National Center for Atmospheric Research, Research Applications Laboratory, Boulder, Colorado. [Available online at: http://dx.doi.org/10.5065/D6WS8R93] Accessed 16 Dec 2015.