Advancements in Objective Multisatellite Tropical Cyclone Center Fixing*

ANTHONY J. WIMMERS AND CHRISTOPHER S. VELDEN

*Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison,
Madison, Wisconsin*

(Manuscript received 2 April 2015, in final form 7 October 2015)

ABSTRACT

An improved version of the Automated Rotational Center Hurricane Eye Retrieval (ARCHER) tropical cyclone (TC) center-fixing algorithm, introduced here as “ARCHER-2,” is presented with a characterization of its accuracy and precision and a comparison with alternative methods. The algorithm is calibrated for 37- and 85–92-GHz microwave imagers; geostationary imagery at visible, near-infrared, and longwave infrared window channels; and scatterometer ambiguities. In addition to a center fix, ARCHER-2 produces a quantitative estimate of expected error that can be used automatically or manually to evaluate the suitability of a result. The median center-fix error ranges from 24 (using scatterometer) to 49 (using infrared window) km relative to the National Hurricane Center best track. Multisatellite, multisensor results can also be used together to produce a TC-track estimate that selects from the best of all of the available imagery in the ancillary “ARCHER-Track” product. The median error of ARCHER-Track varies between 17 and 38 km, depending on TC intensity and data latency. The bias of the product’s expected error varies between 0% and 12%, which translates to an average of only 4 km. When compared with operational, subjective center-fix estimates, the ARCHER-Track approach improves on 29%–43% of these cases at the tropical-depression and tropical-storm stages, at which further assistance is typically sought. This result demonstrates that ARCHER-2 and ARCHER-Track can complement and accelerate operational forecasting where needed and can furnish other automated TC-analysis methods with well-characterized center-fix information.

1. Introduction

Automated center fixing of tropical cyclones (TC) remains an essential component of applications such as objective TC-intensity estimation (Velden et al. 2006; Olander and Velden 2007), detection of rapid intensification (Jiang et al. 2014; Rozoff et al. 2015), TC-structure retrieval (Sitkowski et al. 2011), TC visualization (Wimmers and Velden 2007), and TC climatological descriptions (Knapp et al. 2010; Kossin et al. 2013). Because it is the initial step in most TC-retrieval algorithms, the overall accuracy rests heavily on the accuracy of the center fix. Thus, the need exists for improvement and

sophistication in center fixing, wherever significant innovations can be found.

This work follows from a seminal study that introduced a versatile, objective TC center-fixing algorithm called the Automated Rotational Center Hurricane Eye Retrieval (ARCHER; Wimmers and Velden 2010). Primarily designed for 85–92-GHz microwave imagery, ARCHER allows easier automated access to applications of microwave-based TC diagnostics and visualization. Requiring no manual intervention, ARCHER is well suited for real-time use as well as for analysis of large retrospectively processed TC datasets. ARCHER fixes can also provide an alternative to operational “working” TC-track positions (in real time) and to “best track” positions for retrospective cases. Real-time track forecasts have obvious limitations in accuracy, and 6-h best-track positions can have inaccuracies that arise from track smoothing or from poorly defined rotational centers. The best track can also be suboptimal for an application that requires a center fix that is specific to the retrieved field (i.e., structure relative) and/or navigational offsets of a particular image (i.e., parallax). This

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JAMC-D-15-0098.s1>.

Corresponding author address: Anthony J. Wimmers, Cooperative Institute of Meteorological Satellite Studies, University of Wisconsin–Madison, 1225 W. Dayton St., Madison, WI 53706.
E-mail: wimmers@ssec.wisc.edu

algorithm and its earlier prototypes proved to be useful to algorithms that operate in real time [the automated Dvorak technique (ADT); Olander and Velden 2007] as well as for research purposes (Cossuth 2014; Rozoff et al. 2015).

In this paper, we describe an advanced version of ARCHER, introduced here as “ARCHER-2,” that is applicable to a more complete variety of the satellite-based imagery that is used in TC diagnostics, including geostationary channels, 37-GHz microwave imagery, and scatterometer-derived ambiguity data. ARCHER-2 not only enables improvements in TC-related applications that use these kinds of imagery but also automatically factors each source’s relative merits to provide center-fix confidence information when different sources are used together. This paper introduces ARCHER-2 through a brief review of the legacy algorithm (section 2); a description of the innovations in this latest algorithm (section 3); a calibration of the algorithm expected error (section 4); a validation of the algorithm accuracy, expected-error bias, and comparison with existing alternatives (section 5); and concluding remarks on opportunities for applying the algorithm to research and operations (section 6).

2. Elements of ARCHER

ARCHER was developed to objectively analyze satellite imagery to determine the rotational center of a TC. The original version of ARCHER and ARCHER-2 share the same general operating components. In this section, a brief synopsis of these common components is presented, and then the ARCHER-2 innovations are covered in section 3. Figure 1 is shown for visual reference as an example of how the components outlined in this section operate. For more detailed formulas of the basic algorithm logic and pattern-recognition process, refer to the original description in Wimmers and Velden (2010). To provide a full account of the new algorithm operations and calibrations, the portable programming code (which uses the Matlab software) is publicly available as an online supplement to this paper (<http://dx.doi.org/10.1175/JAMC-D-15-0098.s1>) in the form of a compressed file archive. This package of files includes a training module with examples for testing and evaluation to facilitate adoption by new users.

a. Preprocessing

All input satellite imagery is preprocessed into a new navigation (longitude–latitude grid) that attempts to minimize the effects of parallax. In the absence of rapid, accurate height assignments of TC features that are valid for each sensor, a uniform “feature height” is

assumed throughout the image (Table 1). This assumption is not exact, but it reduces the effects of parallax to a lower-order residual error and greatly improves the precision of intersatellite/sensor comparisons.

A “first guess” position is required input to identify the relevant satellite data and establish the domain of the calculations. In real time, ARCHER uses a short-term forecast track, interpolated to the time of the satellite image, from an applicable operational forecast center [e.g., the National Hurricane Center (NHC) for North Atlantic and eastern Pacific Ocean basins and the Joint Typhoon Warning Center for other basins]. For retrospective processing, an interpolation from the analyzed best track can be used.

b. Spiral score

The spiral score (Fig. 1a) measures how well the gradients in the image align with a spiral configuration that is representative of overall TC structure (i.e., curving rainbands). ARCHER computes a gridded field of scalar values that vary according to the relative alignment between the image gradients and a spiral unit-vector field centered on each corresponding grid point. The spiral-score field normally forms a smooth “bull’s eye” pattern around the maximum score—the best fit of the spiral—leading to a first analysis of the most likely TC rotational center.

c. Ring score

The ring score (Fig. 1b) measures the best fit of the gradients of an inner eyewall (if it exists) to the shape of a circle. Each score in the gridded field is the average dot product of a ring of radially oriented unit vectors and the collocated image gradients on the ring. The result is a gridded field (like the spiral-score field) of the maximum score from a range of ring sizes at each grid point.

d. Distance penalty

The final gridded component is a distance penalty, which varies as the square of the distance from the first-guess point. It is scaled to be a low-order term that is only influential enough to prevent the algorithm from favoring areas toward the edges of the analysis domain, where patterns outside the expected TC-core region can sometimes register false maxima with the spiral- and ring-seeking components, especially in poorly organized TCs.

e. Combined score

The combined-score grid (Fig. 1c) is a weighted sum of the spiral-score grid, the ring-score grid, and the

FIG. 1. Displayed stages of the ARCHER/ARCHER-2 algorithm, for a *GOES-13* image of Hurricane Edouard (category 1) at 0545 UTC 18 Sep 2014: (a) spiral score (evaluating the relative alignment between the image gradients and a spiral pattern), (b) ring score (evaluating the relative alignment between the image gradients and an eyelike ring pattern), (c) combined score (weighted average of the spiral-score field, ring-score field, and a minor penalty for distance from the first guess), and (d) display of the final results, including ARCHER center fix (black square), estimated eye-pattern ring (magenta circle), area of 50% certainty of center-fix location (dashed white circle), and first-guess position for reference (white plus sign).

distance-penalty grid. The objectively determined ARCHER center fix then corresponds to the point on that grid with the maximum combined-score value. A comparison of this final center fix and the first-guess position, without obstructions of a contour overlay, is shown in Fig. 1d.

3. Innovations in ARCHER-2

This section outlines the major upgrades to the algorithm that allow advanced science-application and expected-error comparisons. The innovations are significant

TABLE 1. Characteristics of satellite sensors used in ARCHER-2. In the sensor column, H indicates horizontal polarization.

Sensor	Satellite platforms ^a	Distinguishing features	Feature-height estimate (km)
85–92 GHz (H)	SSM/I (<i>F15</i>), SSMIS (<i>F16</i> , <i>F17</i> , and <i>F18</i>), TMI, <i>GCOM-W1</i> AMSR2, ^b and GMI ^b	Ice scattering from high convection, some liquid droplet scattering	10
37 GHz (H)	TMI ^b , <i>GCOM-W1</i> AMSR2 ^b , GMI ^b , and WindSat ^b	Liquid precipitation	3
IR	GOES-East and other geostationary platforms ^b	Cloud-top temperature	16
SWIR	GOES-East and other geostationary platforms ^b	Cloud-top temperature and sunlight scatter	16
Visible	GOES-East and other geostationary platforms ^b	Sunlight reflectivity	2
ASCAT	<i>MetOp-A</i> ; <i>MetOp-B</i> ^b	Surface wind from ocean wave orientation	0

^a SSM/I: Special Sensor Microwave Imager, SSMIS: Special Sensor Microwave Imager/Sounder, TMI: Tropical Rainfall Measuring Mission Microwave Imager, *GCOM-W1* AMSR2: *Global Change Observation Mission—Water Satellite 1* Advanced Microwave Scanning Radiometer 2, GMI: Global Precipitation Measurement Microwave Imager, and GOES: Geostationary Operational Environmental Satellite.

^b Used operationally in ARCHER-2 but not available in calibration or validation.

enough to warrant the ARCHER-2 label and are summarized in this section.

a. Single-spiral computation

ARCHER-2 uses one, fine-resolution spiral-score calculation (at 0.05° for microwave and 0.025° for geostationary imagery), rather than a coarse-resolution calculation followed by a second, fine-resolution step. We have found the performance and calculation time to be roughly the same for either method, but the ARCHER-2 method has greater conceptual and computational simplicity.

b. Application to multiple channels

Although the original ARCHER was demonstrated on several forms of imagery besides 85–92 GHz, it was not rigorously calibrated for each channel. By contrast, variants of ARCHER-2 can apply to 85–92-GHz microwave imagery as well as 37 GHz and three common geostationary channels: longwave infrared (IR), shortwave infrared (SWIR), and the visible band (Vis). The distinguishing features of each of these variants are organized in Table 1. Note that 37-GHz inputs from SSM/I and SSMIS are not used here, because their lower resolution has been found to yield less-reliable center-fix information.

c. Convective-cell masking

In weaker TCs, an image is sometimes dominated by one or more convective cells that do not follow the surrounding spiral patterns of rotation. In IR imagery, this phenomenon commonly occurs with a large cold cirrus cap; in 85–92-GHz imagery, it commonly occurs with the signature of a high-level ice plume decoupled

from the lower-level circulation. These (often asymmetric) patterns can mislead the center-fix algorithm, and therefore the algorithm runs two complete iterations for these two kinds of imagery. The first iteration proceeds as normal, and the second iteration masks the higher-altitude features with brightness-temperature thresholds of <265 and <245 K for IR and 85–92-GHz imagery, respectively. If at least 50% of the image remains after masking, and the combined score of the masked image exceeds the nonmasked score, then the results for the masked image are used instead.

d. No-fix conditions

When the highest combined score on a gridded field occurs at the edge of the domain, it invariably suggests that an absolute center of a rotational pattern cannot be determined with confidence in that image. In these situations, a “no fix achievable” output is given. However, coincident center-fix output from other sensors will usually compensate for this missing value so as to resolve an effectively continuous storm track, which will be discussed in section 5.

e. Scatterometer-ambiguity module

An additional innovation of ARCHER-2 is the application to the ambiguity vectors of scatterometer retrievals. As described by Figa-Saldaña et al. (2002) in reference to the Advanced Scatterometer (ASCAT) instrument, the ambiguity-vector field is usually the result of from two to four possible solutions to the wind field using the measured radar backscatter of gravity waves on the ocean surface. ARCHER-2 currently limits the application to ASCAT 12.5-km ambiguity retrievals because of its higher directional accuracy, but

FIG. 2. ARCHER-2 diagnostic image for an ASCAT ambiguity retrieval for Tropical Storm Emily (1416 UTC 2 Aug 2011), including the following components: ambiguity vectors (black arrows), unused four-direction ambiguity vectors (red arrows), combined-score field (colored contours), ARCHER center fix (black square), first-guess position (green plus sign), area of 50% certainty of center fix (dashed magenta circle), and area of 95% certainty of center fix (dotted magenta circle).

work is under way to apply the algorithm to other scatterometer instruments as well. Although this module is too different to be described as one more variant of the original algorithm, we include it in the ARCHER-2 package to offer a more complete objective scheme for analyzing satellite-based information that is relevant to TC center fixing.

This module operates in a way that is similar to that of the spiral-score algorithm, and the output is a contoured grid of scores just as in the spiral-score field (Fig. 2). The approach is also similar to the method adopted by expert forecasters: to survey the vector field for the dominant circulation on the basis of the pattern from one set of the ambiguities. In the ARCHER module, the maximum gridpoint value indicates where the ambiguity vectors show the strongest alignment with a spiral pattern centered at that point. All ambiguity vectors are used, and

the vectors that align with the dominant pattern serve to amplify the score of the ultimate center-fix location. One unique feature is an allowance for center-fix points $\leq 0.5^\circ$ (55 km) outside the data region, provided that the contours close around that point. This is necessary with ASCAT data because ASCAT's relatively thin data swath often leads to the center of rotation falling just off the edge of the swath, even when a rotational pattern is well defined. Another unique feature is the removal of any scatterometer retrieval point with four ambiguities, shown in red in Fig. 2. (From two to three ambiguity vectors is more common.) Experience shows that four ambiguity-vector points tend to have unreliable surface wind estimations for this purpose, regardless of the quality-check information, and also tend to occur close to the center of rotation where the center-fixing algorithm has a higher sensitivity to scatterometer error.

FIG. 3. Gamma distribution calculated for $\alpha = 1$. (The PDF formula used by ARCHER-2 and shown here is a common case in which the shape parameter k of the gamma function is 2.)

Moreover, standard quality-check information is not generally helpful in filtering the data for center fixing, because the quality check often flags data for accuracy in vector magnitude alone, which is not as relevant here as vector direction.

f. Expected error

The previous version of ARCHER used an empirically determined threshold value associated with the final score to estimate whether a center fix was likely to be better or worse than an operational forecast/analysis point valid at the same time. ARCHER-2 incorporates a more rigorous calculation of the expected error so that every center fix is accompanied by a quantitative characterization of the center-fix certainty. The expected error is stated in terms of a probability density function (PDF) that follows a gamma distribution (Fig. 3), described by

$$\text{PDF}(x) = \alpha^2 x e^{-\alpha x}, \quad (1)$$

where α is the only parameter necessary to characterize the distribution of the error x . In the next section, we will show how the ARCHER center-fix scoring can be calibrated robustly to this PDF so as to attach any center fix with a corresponding expected-error distribution.

The probability of the true center fix being a distance x from the ARCHER-2 center fix is then given by the cumulative density function (CDF):

$$\text{CDF}(x) = \int_0^x \text{PDF}(x') dx' = 1 - (\alpha x + 1)e^{-\alpha x}. \quad (2)$$

We have observed that the accuracy of the ARCHER-2 center fix depends less on the value of the combined score than on the relative difference between the combined score and the scores of the neighboring points. In other words, the best indication of accuracy is the spacing of the combined-score contours around the center-fix point: the tighter the contour spacing is, the lower is the expected error. This pattern is quantified here as the “confidence score,” defined as the difference between the maximum combined score and the highest score $\geq 0.75^\circ$ away from the location of the image’s combined score (Fig. 4). (This approach is preferable to using a single gradient or Laplacian because it avoids the influence of highly localized variations in the combined score.) Thus, the task of calibrating the expected error of an ARCHER-2 center fix is to determine the error distribution parameter α as a function of the confidence score η , which is described in the next section.

4. Calibration

The calibration dataset for the ARCHER-2 algorithm covers TCs of the North Atlantic basin from 2006 to 2011. The NHC best track (Rappaport et al. 2009) is used for the “truth” center positions, interpolated to the time of each analyzed satellite image. Likewise, simulated real-time first-guess positions are obtained from time-interpolated points along the corresponding archived NHC operational forecast track (from <http://www.nhc.noaa.gov/archive/dis/>), normally at 6-hour updates. Here, the proper NHC operational forecast track is selected while assuming a polar-orbiting-satellite-image latency of 2 h, a geostationary-image latency of 15 min, and a forecast-track-availability latency of 15 min from the track-initialization time.

The following constraints are applied to the ARCHER-2 calibration dataset. The data are limited to analysis times within 3 h of an aircraft reconnaissance fix to ensure the most accurate best-track positions. The dataset is further limited to positions over water, because satellite-derived rotational patterns are often less consistent over land. Analyses poleward of 40° latitude are not considered because TCs undergoing extratropical transition tend to dominate the Atlantic best-track archive in that zone and often have distorted satellite signatures.

Satellite sources for the ARCHER-2 calibration dataset are the same as in Table 1, with one unfortunate exception. The 37-GHz imagery was not available in a readily accessible local archive because it was not introduced into operations until 2014, and therefore it was not included in the calibration/validation. (The local archive is our evolving dataset of global tropical-cyclone

FIG. 4. Examples of the confidence-score calculation in visible-channel imagery for (a) a low-confidence center fix and (b) a high-confidence center-fix. Blue contours are the combined-score fields (contour increment of 0.50), and the blue square is the location of the maximum value, which becomes the center fix. The difference in combined-score values between A and B, explained in the text, is a good quantitative measure of the relative strength of the center-fix point. This difference is called the confidence score.

satellite imagery saved from real-time tests and applications of ARCHER since the late 2000s.) The consequences of this omission are discussed in section 5.

The objective of the calibration process is to fit the shape of the expected error PDF to the ARCHER-2 confidence score η of an image:

$$\alpha(\eta) \equiv m_{\text{sensor}} \eta + b_{\text{sensor}}, \quad (3a)$$

where m_{sensor} and b_{sensor} are the sensor-specific slope and offset, respectively. The fit is improved further by introducing an additional dependency on the forecast maximum sustained wind V_{max} because of a clear transition in storm organization between three regimes of V_{max} : values below 65 kt ($1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$; these values are labeled “lo”), values between 65 and 85 kt, and values above 85 kt (“hi”):

$$\begin{aligned} \alpha(\eta) &\equiv m_{\text{sensor}}^{\text{lo}} \eta + b_{\text{sensor}}^{\text{lo}} \quad \text{for } V_{\text{max}} < 65 \text{ kt} \quad \text{and} \\ \alpha(\eta) &\equiv m_{\text{sensor}}^{\text{hi}} \eta + b_{\text{sensor}}^{\text{hi}} \quad \text{for } V_{\text{max}} > 85 \text{ kt}, \end{aligned} \quad (3b)$$

and $\alpha(\eta)$ is the weighted average of the two values for intensities between 65 and 85 kt.

To evaluate the approximate error distribution for any sensor as a function of η , the results are grouped into quartiles of η and the error distribution is treated as corresponding to the average η within that quartile (Fig. 5). This approach is a fair compromise between the

need to have a large-enough sample size and a small-enough variance in η within each quartile.

The first column of Fig. 5 illustrates the fit of the gamma distribution to the error of the 85–92-GHz center-fix retrievals. (The parameters of the fit between α and η for each sensor, including 85–92 GHz, are provided in Table 2.) The sample in Fig. 5 is limited to the tropical-depression (TD)–tropical-storm (TS) intensities to show the performance at the most difficult, and the most necessary, range of TCs to estimate centers of rotation. It is apparent here that the linear fit using Eq. (3) is appropriate. Note, however, that the empirical fit to the top quartile (top-left histogram) is skewed noticeably toward lower errors than are seen for the actual histogram. This is consistent with the expected difference between image-specific center fixes and the best track. As was observed in Wimmers and Velden (2010, their appendix B), the best track differs from the “true” image-specific center of rotation by roughly 0.15° on average, which means that the actual distribution of error should skew by approximately this amount toward lower values than those of the measured error with respect to the best track.

The remaining three columns of Fig. 5 show that the error of an ARCHER-2 center fix can also be calibrated forward (or backward) in time. This is very important to know if, for example, a low-Earth-orbiting-satellite center fix is to be evaluated against an operational

FIG. 5. Calibration of the expected-error distribution (green line) to the error relative to the best track (blue histogram) for ARCHER-2 applied to 85–92-GHz imagery on tropical cyclones at 35–64-kt intensities (TD–TS). Rows are organized by the four quartiles of confidence-score values. The first column applies to the original imagery at the time of observation. The remaining columns apply to the same center fixes, projected forward by Δt . The parametric function that describes the best fit to the error is given in Eqs. (1) and (3), with parametric values given in Table 2.

center fix at a standard time interval. The method of extending a center fix forward (or backward) in time is to use the equivalent displacement in the available operational forecast track:

$$p_{\text{ext}}(t) = p_{\text{ARCHER}}(t_0) + [p_{\text{op}}(t) - p_{\text{op}}(t_0)], \quad (4)$$

where p_{ext} is the extended position, p_{ARCHER} is the original center fix, p_{op} is the position of the operational forecast track, and the term in square brackets is the displacement. In this study, the operational forecast-track position $p_{\text{op}}(t)$ is the NHC forecast track valid at time t . We recognize that introducing an operational track to the system is a departure from the independent nature of ARCHER-2, but in this case it is a valid design choice for the following reasons. First, by using the

displacement *along* the forecast track rather than the location of the track itself, this method does not cause the ARCHER-2 extended center fix to “gravitate” toward the track; rather, it maintains the same distance

TABLE 2. Parametric fit of ARCHER-2 expected error to confidence score.

	m_{lo}	m_{hi}	b_{lo}	b_{hi}
85–92 GHz	3.28	1.61	2.63	9.58
37 GHz ^a	3.28	1.61	2.63	9.58
IR	9.89	9.26	−2.07	1.95
SWIR	8.68	14.2	−0.37	−0.24
Visible	14.4	14.6	−0.83	3.45
ASCAT	2.67	−19.4	2.67	−19.4

^a Preliminary.

from the track at all times. Second, this approach is far better than using an extrapolation of a history of ARCHER-2 center fixes, because an extrapolation can magnify any small errors over short times. Third, this is the only viable way to account for past, present, and future curvature in the track, often in cases in which the curvature is a foregone conclusion but it would still be difficult to estimate from a higher-order extrapolation of the ARCHER-2 track history.

We find that a robust relationship exists, independent of sensor, between the increase in the spread of the error PDF (corresponding to a decrease in α) and forecast/hindcast time Δt :

$$\alpha(\Delta t) \equiv \frac{\alpha(0)}{1 + (0.00514 \text{ h}^{-1.5})\alpha(0)\Delta t^{1.5}}, \quad (5)$$

where h indicates hours to keep the denominator unitless. This is the formula applied to the remaining three columns of Fig. 5. According to this formula, the spread of error in an extended center fix increases exponentially with the extension out in time. This relationship is very useful, because in certain applications it is often necessary to know whether to rely on a high-confidence center fix extended over a certain length of time or a lower-confidence center fix extended over a shorter time, as will be explored in the following section.

5. Validation

The ARCHER-2 algorithm is validated in two ways: first, by characterizing the error of each sensor-specific center-fixing component and, second, by evaluating the ability of a multisensor blended product to provide TC-track guidance. The validation dataset is independent of the calibration dataset and covers the 2012 Atlantic hurricane season, filtered for latitude, land, and data availability (in the local archive) as with the calibration dataset. One key difference, however, is that the dataset is not filtered for times near aircraft reconnaissance, because with such breaks in data continuity we could not evaluate the product's ability to represent a storm track. Also, TCs Oscar (2–5 October) and Sandy (21–31 October) were not included because of an unfortunate loss of data in the local file system, and to recover every occurrence of missing data would be time prohibitive. The remaining data still provide a comprehensive distribution of TC stages and intensities for evaluation purposes.

Table 3 summarizes the center-fix error validated for each sensor, subdivided by TC intensity. The validation statistics used here are the median error (directly

TABLE 3. Validation of ARCHER-2, by sensor type and TC intensity. Here and in subsequent tables, ME indicates median error and N is sample size.

		ME (km)	Bias (%)	No fix (%)	N
85–92 GHz	TD–TS	35	−0.6	11.2	632
85–92 GHz	Category 1	26	−11.2	11.0	127
85–92 GHz	Categories 2–5	11	+7.6	2.2	45
85–92 GHz	All	31	−1.8	10.7	804
IR	TD–TS	56	+23.8	38.2	570
IR	Category 1	43	+7.9	17.3	110
IR	Categories 2–5	21	−3.0	14.3	49
IR	All	49	+18.5	33.5	729
SWIR	TD–TS	46	+20.9	32.8	570
SWIR	Category 1	41	+9.4	16.4	110
SWIR	Categories 2–5	21	+0.3	12.2	49
SWIR	All	43	+17.2	28.9	729
Vis	TD–TS	36	+18.0	35.3	278
Vis	Category 1	26	+0.7	19.3	57
Vis	Categories 2–5	22	−7.4	24.0	25
Vis	All	32	+12.8	31.9	360
ASCAT	TD–TS	26	−8.9	0.0	49
ASCAT	Category 1	18	−8.4	0.0	12
ASCAT	Categories 2–5	16	+9.6	0.0	4
ASCAT	All	24	−7.6	0.0	65

comparable to the 50th percentile of the expected error), the bias of the expected error (explained shortly), and the percentage of no-fix analyses (in which the algorithm could not generate a confident center fix). It can be seen that 85–92-GHz and ASCAT center fixes have the most reliable results overall, with a low percentage of no-fix cases and a lower error than that from geostationary channels. ASCAT center fixes are more often the most accurate for TD–TS and category-1 intensities, and 85–92-GHz center fixes are most often the most accurate for categories 2–5. Of the three geostationary channels, the visible channel is the most accurate at intensities ranging from TD to category 1, but the channels are about equal at higher intensities. SWIR has lower error than IR at TD–TS intensities (although still high relative to that of the other sensors), but the differences are almost negligible for storms in categories 1–5.

The relative merits of each sensor are clearly apparent in these results. Microwave 85–92-GHz images yield a low rate of no-fix results, the second-lowest error for weaker TCs, and a sampling rate that is close to three per hour. These sensors consequently tend to dominate in a combined-sensor scheme. Geostationary imagery (IR, SWIR, and visible) have the highest average fix errors, but can “fill in the gaps” between polar-satellite passes when there is a detectable rotational pattern in the imagery. Of the geostationary channels, the visible channel performs almost as well as

TABLE 4. Validation of ARCHER-2, by sensor type and expected error statistic α .

	α	ME (km)	Bias (%)	<i>N</i>
85–92 GHz	9.0– ∞	31	–20.3	53
85–92 GHz	6.0–9.0	29	–3.7	189
85–92 GHz	3.0–6.0	39	+3.8	306
85–92 GHz	0.0–3.0	34	+21.6	13
IR	9.0– ∞	18	–10.3	3
IR	6.0–9.0	30	–8.9	26
IR	3.0–6.0	41	+1.3	65
IR	0.0–3.0	66	+33.2	258
SWIR	9.0– ∞	17	–6.8	4
SWIR	6.0–9.0	28	–4.1	24
SWIR	3.0–6.0	34	+8.8	97
SWIR	0.0–3.0	59	+28.3	258
Visible	9.0– ∞	26	–12.4	7
Visible	6.0–9.0	24	+6.2	31
Visible	3.0–6.0	27	+9.4	50
Visible	0.0–3.0	55	+29.0	92
ASCAT	9.0– ∞	24	–12.7	33
ASCAT	6.0–9.0	37	–12.2	9
ASCAT	3.0–6.0	23	+14.4	3
ASCAT	0.0–3.0	70	+12.9	4

the microwave imagery in the range of TD–category 1 but is limited to daylight hours. The ASCAT center fixes are a highly accurate and valuable resource for the range from TD to category 1, but the orbits and narrow swath width limit its availability to an infrequent number of events, as shown in the low sample number.

With regard to expected-error bias, in this study the bias is defined as the average offset of the expected error from the actual error, expressed in terms of a percentile difference. For example, in a single application of ARCHER-2, if the measured error corresponds to the 60th percentile on the expected-error distribution, then the bias of that estimate is –10%. In this scheme, negative values of bias mean that the algorithm's expected error understated the measured error and positive values indicate an overstatement. For each sensor, the bias is within 1%–11% of a perfect calibration for intensities in categories 1–5. For errors in the range of 25 km, a bias of 11% equates to a 4-km difference in expectations, which is close to negligible in a forecasting setting. On the other hand, the somewhat larger biases for cases in the TD–TS range warrant further examination.

Table 4 gives an alternate breakdown of bias according to derived α values, which are in four nearly equally spaced bins of expected error. The highest α range corresponds to the highest-confidence center fixes, and the lowest α range corresponds to the lowest-confidence center fixes. The common trend for each sensor is that

some of the highest-magnitude biases in expected error come from the results with the highest-confidence center fixes. To visualize an example of this error distribution and its bias, the error of the 85–92-GHz dataset is shown in the first column of Fig. 6. The same pattern as was seen in Fig. 5 from the calibration is evident in the top-left histogram in Fig. 6: the fit of the expected error is skewed toward slightly lower values, which is most evident when the distribution of the error is the narrowest. Inspection of the individual images from this group confirms that the expected error fit (green line in the top-left histogram) is indeed a more accurate representation of the error apparent in the imagery, because of the difference between the best track and the true image-relative center of rotation. The high-magnitude bias values in the lowest quartiles in Table 4 are not a major concern, because in the operational application of this algorithm the center fixes with low confidence will be rejected in favor of higher-confidence center fixes from other sensors that are close in time.

This leads to the second part of the validation: How well does the algorithm work as a complementary, multisensor composite approach to creating a continuous TC track? For the purposes of segregating this particular application of ARCHER-2, the approach is dubbed “ARCHER-Track.” For effective visualization, the composite track provides a center fix at 3-h intervals, which necessitates a conversion of polar-satellite fixes to the evenly spaced ARCHER-Track time intervals. The remaining three columns in Fig. 6 demonstrate that the method of center-fix extensions [Eq. (4)] is well characterized for all but the lowest-certainty center fixes (bottom row), which are also the least likely to be used in a composite. Furthermore, the method of extending center fixes in time is usually limited to differences of ≤ 3 h, and therefore the effect is generally small. The four columns of 0, 3, 6, and 9 h simply reveal that the approximation in Eq. (4) is robust.

The method of ARCHER-Track is to use the highest-confidence center fix from the collection of available fixes at each time step. If there are no quality observations over a 3-h time interval, a temporal extension or interpolation can be used, which is explained as follows. In a real-time application, a lack of confident center fixes over the *latest* time interval will require a temporal extension of a previous result (usually extending ≤ 3 h), whereas time gaps for *retrospective* processing of tracks can employ interpolation. For the 2012 season of TCs validated here, we use a maximum interpolation gap of 12 h, but this threshold should be adjusted each season according to the changing availability of satellite sources for center fixing. Since 2012, the number of sensors in use is higher, and consequently the maximum interpolation gap is decreased to 6 h.

FIG. 6. Validation of the expected-error distribution (green line) to the error relative to the best track (blue histogram) for ARCHER-2 applied to 85–92-GHz imagery on tropical cyclones at 35–64-kt intensities (TD–TS), as in Fig. 5. The green line is the expected error from ARCHER-2, with the bias between the expected error and the actual best-track error given in Table 4.

Figure 7 presents an example of the ARCHER-Track product over the lifetime of Hurricane Michael (2012) and shows the typical pattern of how the various satellite sensors are used. The display is designed to make the center fixes and their expected error intuitively clear: fix positions are the colored dots, the 50% confidence range is indicated by the surrounding circle, and the lightly shaded region is the moving 95% confidence range. The colors correspond to the satellite sensor used for the fix, designated in the image legend. In Fig. 7, the underlying white disk is the NHC best track interpolated to 3-h time steps and marked with the TC intensity. (In real time, however, the underlying disks would be the operational analysis and forecast track.)

This case is typical in the way that microwave imagery and ASCAT are predominantly selected in the TD–TS intensities, and then the geostationary sensors become

better represented as intensity increases. Somewhat surprising is that the SWIR fixes are better represented than IR in this case. Upon further inspection, the results from SWIR and IR are almost equivalent during nighttime, and SWIR center fixes from GOES platforms have a slightly lower expected error than does IR, and IR center fixes have a slightly lower expected error than SWIR when using geostationary satellites.

On occasion, as the selection shifts from one satellite sensor to another, there is also a shift in offset. This might be due to parallax residuals, vortex tilt, or systematic location bias. This effect is especially apparent in the change from geostationary data to microwave data in the last westward leg of the storm track. Basic algorithm error can play at least as large a role in creating these shifts, however. For example, the discrepancy between the two sequential visible-channel fixes at 32.5°N, 42.0°W

FIG. 7. Example of the ARCHER-Track product for Hurricane Michael (2012). Components of the graphic are explained in the top-right legend. Inside the white circles, D = tropical depression, S = tropical storm, 1 = category-1 hurricane, etc.

shows that not all of the sudden shifts in position are due to changes in selected sensors. It will be the task of the expert user to observe these subtler patterns and arrive at the best estimate of the TC development and motion.

Next, the relative character of ARCHER-Track results during the 2012 North Atlantic season is compared with that of ARCHER-Track run globally in real time with all available satellite platforms during 2014. For the

FIG. 8. Components of the near-real-time track for the 2012 validation in the North Atlantic.

2012 validation dataset (Fig. 8), 85–92-GHz microwave dominates the center fixes for intensities from TD to category 1, and ASCAT plays an especially large role relative to its low sampling rate. At category 2 and above, geostationary imagery is represented in more than one-half of the center fixes, and 85–92 GHz makes up approximately one-third. The makeup of the 2014 ARCHER-Track (Fig. 9) is similar except for the new influence of 37-GHz microwave center fixes. The influence of microwave channels is slightly higher across all TC intensities because of the addition of this channel. At TD-TS intensities, the 37-GHz imagery and the 85–92-GHz imagery are roughly equal in representation, and then the influence of 37 GHz wanes with increasing TC intensity. The higher representation of IR imagery in 2014 for TCs in categories 2–5 is likely due to the slightly higher accuracy (lower estimated errors) of IR relative to SWIR imagery in non-GOES imagers, as discussed earlier.

For potential real-time applications, the automated ARCHER-Track system is designed to continuously

update as new data become available. This means that center fixes from real-time geostationary imagery can be replaced by longer-delayed but higher-confidence polar-orbiter fixes as the images become available. Thus, a comprehensive evaluation of the ARCHER-Track system requires separate validations for the “real time” center fixes and the “near real time” center fixes (which become the final resolved track).

Table 5 characterizes the error of the ARCHER-Track results from the 2012 sample using the NHC best track as validation, under the two scenarios of real time (using only the objective fix information available at the valid time) and near-real-time (using all available objective fix information). Because the real-time category relies very heavily on geostationary data (with occasional contributions from time-extended polar-satellite data), the errors are predictably close to the errors of the geostationary sensors from Table 3. The results for the near-real-time scenario, on the other hand, demonstrate the added benefit of using a complementary array of

FIG. 9. Components of the near-real-time track for 2014 global operations.

TABLE 5. Validation of ARCHER-Track relative to NHC best track. Here, AE indicates average error.

		ME (km)	Bias (%)	AE (km)	N
Real time	TD-TS	38	-11.8	52	464
	Category 1	29	-12.3	30	69
	Categories 2-5	18	-11.1	19	36
	All	33	-11.8	48	569
Near-real time	TD-TS	32	-2.6	42	456
	Category 1	24	-0.7	26	69
	Categories 2-5	17	-13.9	16	36
	All	30	-3.8	39	561

sensor inputs. The median error is lower than that from any single sensor, except for ASCAT, which is not available very often. The improvement in error over the real-time cases for intensities of TD-category 1 is especially significant, meaning that in this intensity range the ARCHER-Track is likely to make noticeable improvements in an estimated TC position as that position goes from 0 to ~3 h old. Note also that the expected error bias is reasonably small in all categories (from 0% to 12% in magnitude) and is negative, meaning that the expected error of the ARCHER-Track composite tracks is highly accurate and tends toward a slight understatement of the error relative to best track, which is a closer estimate to the true error, as discussed earlier.

The final method of validation is to compare the accuracy of ARCHER-Track with a forecaster's alternatives, both to assess the relative skill of ARCHER-Track and to examine the usefulness of this algorithm in TC forecasting/nowcasting. Three center-fix sources from 2012 are employed for an independent comparison, using the NHC best track again for validation. The first

source is the set of TC positions interpolated from the NHC operational forecast tracks available at the valid times of the ARCHER-Track product (NHC fx/an). The remaining two sources are the record of Satellite Analysis Branch (SAB) and the Tropical Analysis and Forecasting Branch (TAFB) manual Dvorak center fixes. The previous filtering conditions are applied here as well: no cases over land, only cases equatorward of 40°N, and limited to times of full data availability in our local archive. Also, because the valid times for ARCHER-Track were calculated at 15 min *after* the hour because of the constraints of our archive, the error of these points relative to the interpolated best track has to be compared with the error of their counterparts 15 min *before* the hour, which is the usual schedule for Dvorak center fixes in the North Atlantic. Because the time offset from the top of the hour is the same in each case, this technique will not bias the results.

Comparison statistics for these three independent sources are listed in Table 6. The error for each of these sources is roughly equivalent and is ~30%–40% lower than that from the ARCHER-Track results of Table 5. This should be expected, given that the center fixes from these sources come from an application of expert knowledge to all of the same satellite data used in ARCHER-Track as well as other nonsatellite data. ARCHER-Track was not designed to substitute for a skilled TC analyst, however. Rather, it serves to complement and accelerate the forecasting process, in part by providing a rapid, high-frequency output capability and also by highlighting information that may not otherwise receive proper notice.

To evaluate the guidance aspect of ARCHER-Track in this regard, the percentages of center fixes with lower error than their operational counterparts are also listed

TABLE 6. Accuracy of ARCHER-Track relative to alternative real-time forecasting methods (defined in the text). The percentage of cases in which real-time ARCHER-2 performed better than the listed source is indicated by %AR RT, and %AR NR is similar but for near-real-time ARCHER-2.

		ME (km)	AE (km)	% AR RT	% AR NR	N
NHC fx/an	TD-TS	25	31	29	37	456
	Category 1	14	17	25	38	69
	Categories 2-5	11	12	25	42	36
	All	22	28	28	37	569
SAB	TD-TS	26	34	31	40	239
	Category 1	20	25	25	42	36
	Categories 2-5	11	12	21	26	19
	All	24	31	30	39	294
TAFB	TD-TS	25	34	37	43	439
	Category 1	20	25	25	44	68
	Categories 2-5	12	13	26	21	36
	All	23	31	35	42	543

in Table 6 (“%AR RT” and “%AR NR”). These percentages are significantly high, ranging from 29% to 43% at TD–TS intensities, where further assistance is typically sought. Note also several other factors that are not included in this analysis, each of which fall in ARCHER’s favor. First, because ARCHER includes an expected-error estimate with every center fix, a forecaster would know in advance which results from ARCHER are likely to be more/less accurate than normal. Second, this comparison does not include the contribution to ARCHER-Track from 37-GHz imagery, which improves its performance further, as shown in the 2014 trials. Third, this also does not include the contribution of the AMSR2, on board the *GCOM-W1*, or the GMI, which performs with low-enough image latency to make significant improvements to the *real-time* performance of ARCHER-Track as well as the near-real-time performance.

6. Conclusions

The improved capabilities of ARCHER-2 open up new possibilities for TC applications. ARCHER-2 can serve as an important TC-analysis aid with its rapid automated center-fix retrievals from complementary, multiplatform satellite sensors. Much of the center-fix and expected-error information can be incorporated into visualizations that not only present the important information intuitively but also allow users to evaluate and compare the results with a new level of precision. This includes the interrogation and intercomparison of center-fix information from several satellites within a time window, or simply establishing the appropriate level of confidence for the center fix from a given satellite source. A real-time ARCHER-2 Internet site (<http://tropic.ssec.wisc.edu/real-time/archerOnline/web/index.shtml>) has been established as a test bed for forecast-assisting techniques. The algorithm is currently being operationally assessed by NHC as part of the NOAA Joint Hurricane Testbed program.

The enhanced capabilities of ARCHER-2 also benefit other automated TC-analysis applications. The ADT (Olander and Velden 2007) and the TC-rapid-intensification techniques of Jiang et al. (2014) and Rozoff et al. (2015) require a center fix as the first step, and their performance can improve significantly with even minor improvements to center-fix accuracy. Furthermore, the ancillary expected error associated with every ARCHER-2 center fix provides a consistent and quantitative metric for another algorithm to use to decide whether the ARCHER center fix or an operational-center forecast track is more applicable at a given time.

The added precision and comparability of fix results from various sensors can also lead to an improved

understanding of the meaning of a “rotational center.” In a major TC with a clear eye there is no confusion over this definition, but in weaker storms this is not always so clear because of ambiguities in the concept of what constitutes a storm center. For example, differences in inferred storm center of rotation can occur between Eulerian (fixed location) and Lagrangian (storm relative) frames of reference. In specific terms, aircraft reconnaissance, fixed-sensor measurements, and scatterometer retrievals normally indicate Eulerian rotation whereas all other satellite imagery indicates a Lagrangian center of rotation because of its representation of atmospheric tracers. In practice, these differences are usually overlooked when making a lower-precision consensus estimate of TC position, but, as automated retrievals such as those from ARCHER-2 increase precision, this presents a possible future opportunity to characterize weaker systems more naturally. It could mean distinguishing between the Eulerian and Lagrangian centers, estimating vortex tilt, and quantifying possible asymmetry of the rotation, each of which could improve the process of TC analysis and short-term track forecasting.

Acknowledgments. This work was sponsored by the Oceanographer of the Navy through the PEO C4I PMW-150 program office and the Naval Research Laboratory (Jeff Hawkins). The validation and real-time applications described here are sponsored by the Joint Hurricane Testbed of the National Hurricane Center. We sincerely thank three reviewers for volunteering their time to help to improve this manuscript.

REFERENCES

- Cossuth, J. H., 2014: Exploring a comparative climatology of tropical cyclone core structures. Ph.D. dissertation, Florida State University, 201 pp. [Available online at <http://diginole.lib.fsu.edu/etd/8965/>.]
- Figa-Saldaña, J., J. J. W. Wilson, E. Attema, R. Gelsthorpe, M. R. Drinkwater, and A. Stoffelen, 2002: The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: A follow on for European wind scatterometers. *Can. J. Remote Sens.*, **28**, 404–412, doi:10.5589/m02-035.
- Jiang, H., M. Kieper, and Y. Pei, 2014: Improvement to the satellite-based 37 GHz ring rapid intensification index. Presentation, 68th Interdepartmental Hurricane Conf./Tropical Cyclone Research Forum, College Park, MD, NOAA/Center for Weather and Climate Prediction, paper s02-02. [Available online at <http://www.ofcm.gov/ihc14/presentations/Session2/s02-02jiang.pdf>.]
- Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, doi:10.1175/2009BAMS2755.1.

- Kossin, J. P., T. L. Olander, and K. R. Knapp, 2013: Trend analysis with a new global record of tropical cyclone intensity. *J. Climate*, **26**, 9960–9976, doi:[10.1175/JCLI-D-13-00262.1](https://doi.org/10.1175/JCLI-D-13-00262.1).
- Olander, T. L., and C. S. Velden, 2007: The advanced Dvorak technique: Continued development of an objective scheme to estimate tropical cyclone intensity using geostationary infrared satellite imagery. *Wea. Forecasting*, **22**, 287–298, doi:[10.1175/WAF975.1](https://doi.org/10.1175/WAF975.1).
- Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419, doi:[10.1175/2008WAF2222128.1](https://doi.org/10.1175/2008WAF2222128.1).
- Rozoff, C., C. Velden, J. Kossin, and J. Kaplan, 2015: Improvements in the probabilistic prediction of tropical cyclone rapid intensification with passive microwave observations. *Wea. Forecasting*, **30**, 1016–1038, doi:[10.1175/WAF-D-14-00109.1](https://doi.org/10.1175/WAF-D-14-00109.1).
- Sitkowski, M., J. Kossin, and C. Rozoff, 2011: Intensity and structure changes during hurricane eyewall replacement cycles. *Mon. Wea. Rev.*, **139**, 3829–3847, doi:[10.1175/MWR-D-11-00034.1](https://doi.org/10.1175/MWR-D-11-00034.1).
- Velden, C. S., and Coauthors, 2006: The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bull. Amer. Meteor. Soc.*, **87**, 1195–1210, doi:[10.1175/BAMS-87-9-1195](https://doi.org/10.1175/BAMS-87-9-1195).
- Wimmers, A. J., and C. S. Velden, 2007: MIMIC: A new approach to visualizing satellite microwave imagery of tropical cyclones. *Bull. Amer. Meteor. Soc.*, **88**, 1187–1196, doi:[10.1175/BAMS-88-8-1187](https://doi.org/10.1175/BAMS-88-8-1187).
- , and ———, 2010: Objectively determining the rotational center of tropical cyclones in passive microwave satellite imagery. *J. Appl. Meteor. Climatol.*, **49**, 2013–2034, doi:[10.1175/2010JAMC2490.1](https://doi.org/10.1175/2010JAMC2490.1).