CYCLONE CENTER

Can Citizen Scientists Improve Tropical Cyclone Intensity Records?

BY CHRISTOPHER C. HENNON, KENNETH R. KNAPP, CARL J. SCHRECK III, SCOTT E. STEVENS, JAMES P. KOSSIN, PETER W. THORNE, PAULA A. HENNON, MICHAEL C. KRUK, JARED RENNIE, JEAN-MAURICE GADÉA, MAXIMILIAN STRIEGL, AND IAN CARLEY

Crowdsourcing could help alleviate the uncertainty inherent in the modern global tropical cyclone record.

One of the fundamental challenges in tropical cyclone (TC) analysis and forecasting is accurately determining the storm's sustained maximum wind speed (or "intensity") in an area with little or no in situ observations. Forecast centers, relying primarily on interpretation of geostationary satellite imagery as well as any other available data during operational activities, create postseason "best track" datasets. Best-track data contain (at minimum) information on TC track and intensity. They are widely used in a large number of research applications, including trend analysis (e.g., Emanuel 2005; Webster et al. 2005; Wu et al. 2006), forecast verification (e.g., Sievers et al. 2000; Poroseva et al. 2010), and evaluation of reanalysis datasets (Schenkel and Hart 2012).

Unfortunately, when tropical cyclones are tracked by more than one agency, best-track data frequently disagree. For example, Webster et al. (2005) showed that the frequency of category 4 and 5 typhoons in the northwestern Pacific Ocean increased by 41% between the 1975–89 and 1990–2004 periods. But Wu et al. (2006) reported that those severe storms actually decreased in frequency between 10% and 16% over the same time periods. This contradiction arises from the source of the best-track data used by each group. Webster et al. used data from the Joint Typhoon Warning Center (JTWC); Wu et al. used best-track data

AFFILIATIONS: HENNON—University of North Carolina at Asheville, Asheville, North Carolina; KNAPP—NOAA/National Climatic Data Center, Asheville, North Carolina; SCHRECK, STEVENS, P. A. HENNON, AND RENNIE—Cooperative Institute for Climate and Satellites (CICS-NC), North Carolina State University, Asheville, North Carolina; KOSSIN—NOAA/National Climatic Data Center, Asheville, North Carolina, and NOAA/Cooperative Institute for Meteorological Satellite Studies, Madison, Wisconsin; THORNE*—Nansen Environmental and Remote Sensing Center, Bergen, Norway; KRUK—ERT, Inc. at NOAA/National Climatic Data Center, Asheville, North Carolina; GADÉA—Citizen Scientist, Montpellier, France; STRIEGL—Citizen Scientist, Schortens, Germany; CARLEY—Citizen Scientist, Brisbane, Queensland, Australia

***CURRENT AFFILIATION:** Maynooth University Department of Geography, Maynooth Ireland

CORRESPONDING AUTHOR: Christopher C. Hennon, University of North Carolina at Asheville, 1 University Heights, CPO #2450, Asheville, NC 28804

E-mail: chennon@unca.edu

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-13-00152.1

In final form 18 July 2014

©2015 American Meteorological Society

from the Regional Specialized Meteorological Centre (RSMC) Tokyo and the Hong Kong Observatory. To further complicate matters, Kossin et al. (2007) applied an objective intensity algorithm to homogeneous infrared (IR) satellite data and found no significant change in severe typhoons from 1983 to 2005.

Kossin et al. (2013) provide a good discussion on the causes of differences in global best-track data. They can be roughly classified into two categories: changes in technology and data availability and diverse analysis methods at RSMCs. Aircraft reconnaissance provides forecasters with the highest amount of confidence for a storm's position and intensity. Data are recorded at flight level and GPS dropsondes are routinely deployed during missions to provide surface and vertical-profile information. Missions began during the 1940s in the Atlantic and North Pacific; unfortunately, routine missions stopped in the North Pacific in 1987 and have never been regularly flown in any other ocean basin besides the North Atlantic. Shoemaker et al. (1990), Gray et al. (1991), and Martin and Gray (1993) provide quantitative evidence for the effectiveness of aircraft reconnaissance in reducing TC initial and forecast error. Furthermore, in a recent survey of National Hurricane Center specialists, Landsea and Franklin (2013) found that intensity and position uncertainty decreases substantially when aircraft data are available.

Because reconnaissance data are rarely available [even in the North Atlantic, only about 30% of best-track times include aircraft data Rappaport et al. (2009)], analysts must turn to geostationary satellites, microwave satellites, and sparse in situ measurements such as buoys or nearby ships; it is then that the determination of intensity becomes subject to analyst interpretation and RSMC rules. Wu et al. (2006) attribute differences between JTWC and RSMC Tokyo to wind speed time averaging (10 versus 1 min), the discrete nature of the Saffir–Simpson scale (Kantha 2006), and an "in-house" change to the application of the Dvorak technique at RSMC Tokyo (Koba et al. 1990). The Dvorak technique (Dvorak 1973, 1984; Velden et al. 2006) is a procedure that provides an analyst with an estimate of the maximum sustained winds of a TC based on its cloud pattern, cloud-top temperature considerations, and recent intensity trend. It is used at all RSMCs and is widely considered the best available tool for determining TC intensity in the absence of direct observations. However, the method is also inherently subjective and its rules have been changed at various RSMCs to conform to perceived regional intensity differences. This is discussed in more detail in the section on "Data and design."

The conflicting conclusions regarding western Pacific TC intensity trends, significant differences found in other basins (Schreck et al. 2014), and the changes in technology and observation practices that have led to these discrepancies significantly lower our confidence in the TC record. In fact, recent assessment work by the World Meteorological Organization (WMO) (Knutson et al. 2010) and the Intergovernmental Panel on Climate Change conclude there is "low confidence that any observed long-term (i.e. 40 years or more) increases in TC activity are robust, after accounting for past changes in observing capabilities" (Seneviratne et al. 2012).

This statement and the work summarized above suggest that a comprehensive global reanalysis of TC intensity is needed. There are a number of recent and ongoing projects in this area, but they are restricted to analyses of single storms (e.g., Landsea et al. 2004), specific time periods (e.g., Landsea et al. 2008), or regions (Diamond et al. 2012; T. Kimberlain and P. Caroff 2014, personal communication). Although all of these efforts are valuable, they are not ideal for two reasons. First, they are necessarily restrictive because of the large amount of time required to analyze voluminous amounts of data and imagery. For example, here we are analyzing nearly 300,000 images from a 32-yr TC data record. Assuming 5 min per analysis, it would take one person 25,000 h (~12 years) to analyze all of the images one time—and that is with no time off for holidays or vacation. Second, regional reanalyses risk exacerbating interbasin differences because they are not applied globally or consistently.

We present a new approach to TC reanalysis that is global in scope, can be completed in a reasonable amount of time, and shows promising skill in estimating TC intensity. Cyclone Center is an Internet portal that provides global, homogeneous, TC-centric IR satellite imagery to "citizen scientists." Instead of using a small number of experts, we tap into the scientific curiosity of thousands of ordinary people, using their enthusiasm, time, and pattern recognition skills to eventually answer these scientific questions:

1) Can citizen scientists provide more certainty in TC intensity when forecast agencies disagree?
2) Can citizen scientists provide more skill than current approaches in determining TC intensity, particularly with difficult patterns where automated techniques struggle?
3) Has TC intensity responded to recent changes in climate?

Our intent is not to replace or change best-track data; rather, we intend to provide a global record of TC intensity as determined by a consistent algorithm applied to a homogeneous record of satellite data. Such a data record could then be used as one important tool in any future global TC reanalysis effort. Therefore, this paper will not completely answer the questions above but will describe the ideas and motivation behind Cyclone Center, the developments in data that have made it possible, the process of soliciting and collecting citizen scientist responses, and some of the early results that begin to address these questions. Interested readers can participate in Cyclone Center at http://cyclonecenter.org.

THE CROWDSOURCING OF SCIENCE.

Cyclone Center is one of nearly 30 projects that make up the Zooniverse, a website that started in July 2007 with the launch of Galaxy Zoo (Lintott et al. 2008). With the goal of identifying galaxies for further study, Galaxy Zoo aimed to use the ability and enthusiasm of the general public to perform tasks previously left to trained experts. By obtaining analyses from several volunteers, it became possible to analyze much greater quantities of data and to assign a level of confidence to each classification. The results proved as reliable as those from experts in the field. The project started very simply, asking volunteers only a handful of very simple questions about the shape and orientation of each galaxy, but grew more complex as subsequent revisions were made.

Following the success of Galaxy Zoo, several other astronomy-based projects were developed to investigate a wide variety of celestial imagery and data. In autumn 2010, Old Weather became the Zooniverse's first project based on climate data. The site asks volunteers to help transcribe the daily weather logs kept by ships in the early twentieth century, with the aim of supplementing the historical weather record with the only oceanic data available at the time. With Old Weather's success, it became evident that there was adequate enthusiasm for climate data among many in the general public.

Cyclone Center was a logical next step, since it combined this developing interest in weather and climate with an analysis method that has several parallels with Galaxy Zoo. Both projects require the visual inspection of remotely sensed data with the hope of applying human pattern recognition to the subject in question. They also focus on using human ability to identify spiral patterns in noisy pictures. These parallels meant that some of the web interface code from Galaxy Zoo could be repurposed for Cyclone Center.

ENGAGING THE CROWD. *Crowdsourcing strategies.* Crowdsourcing offers the ability to quickly analyze vast amounts of data. Naturally, that advantage relies on attracting and maintaining a large community of citizen scientists. One might expect that the destructive power of tropical cyclones would be enough to attract and maintain a large number of users, but Cyclone Center still had to be carefully designed to do so. The development team at the Citizen Science Alliance (CSA; the parent organization of Zooniverse) was a valuable resource during this process, since they could call on past experiences from other projects to estimate the abilities of our volunteers.

Citizen scientists, especially nonexperts, hold valid concerns as they participate. One of the most common comments from our volunteers has been "How do I know if I'm doing it right?" This has proven a very difficult question to answer because it is not feasible to analyze their classifications in real time. One of the goals of this paper, as well as others that may follow, is to quantitatively investigate this question. Real-time support is available to participants; many citizen science projects, including Cyclone Center, have active forum communities where volunteers can discuss classifications with each other and even with the science team. Classifiers build their confidence through these discussions, and sometimes they even lead to serendipitous discoveries (Cardamone et al. 2009).

Another way that we built volunteer confidence was through online training. We provide the volunteers with detailed visual guidance that explains each question. In designing this guidance, we had to find ways to provide the necessary information without overwhelming our audience. One of the more successful forms of training has been the tutorial that was recently added to the site, which takes a sample image from Super Typhoon Keith (1997) and gives volunteers a step-by-step description of how to properly classify it.

We also engage and educate our volunteers through Facebook, Twitter, and blog posts. We use these outlets to share preliminary results, conference presentations, and peer-reviewed articles like this one—all of which provides confidence to our volunteers that they really are contributing to science. We also build their interest in tropical cyclones with posts describing recent events like Hurricane Sandy and the start or end of the Atlantic hurricane season. Some of our most viewed posts educate readers on how tropical cyclones form and why they have eyes. Much of that traffic comes from unanticipated Internet searches that direct people to our blog, from which they can discover and join our project.

Citizen science participation. The dataset selected for this study includes roughly 239,000 classifications from just over 5,000 volunteers—basically all of the data collected from the first year of the project. The distribution of users is such that a large majority have completed just a handful of classifications, while a much smaller core of very motivated users have completed orders of magnitude more. Figure 1a shows a breakdown of classifications completed by each user. While the mean number of classifications is just under 50 per user, only 10% have completed more than this, and nearly half have completed just six or fewer.

At the top end of the spectrum, however, a small contingent of "power users" have contributed immense numbers of classifications to the study, with more than a dozen users completing over 1,000 classifications, and a single user even surpassing 10,000—more than the bottom 3,000 users combined. These power users do not strongly bias the data, however—each user is restricted to one classification per image.

Figure 1b shows the contribution of each of these tiers to the total activity on the site. While 28% of users have completed between one and three classifications, their total contribution only accounts for roughly 1% of the total data collected. In contrast, the top 1% of classifiers, those who have completed more than 500 each, account for over 40% of the site's total activity. This highlights the importance of keeping power users engaged, as two-thirds of our collected data have come from just 5% of users.

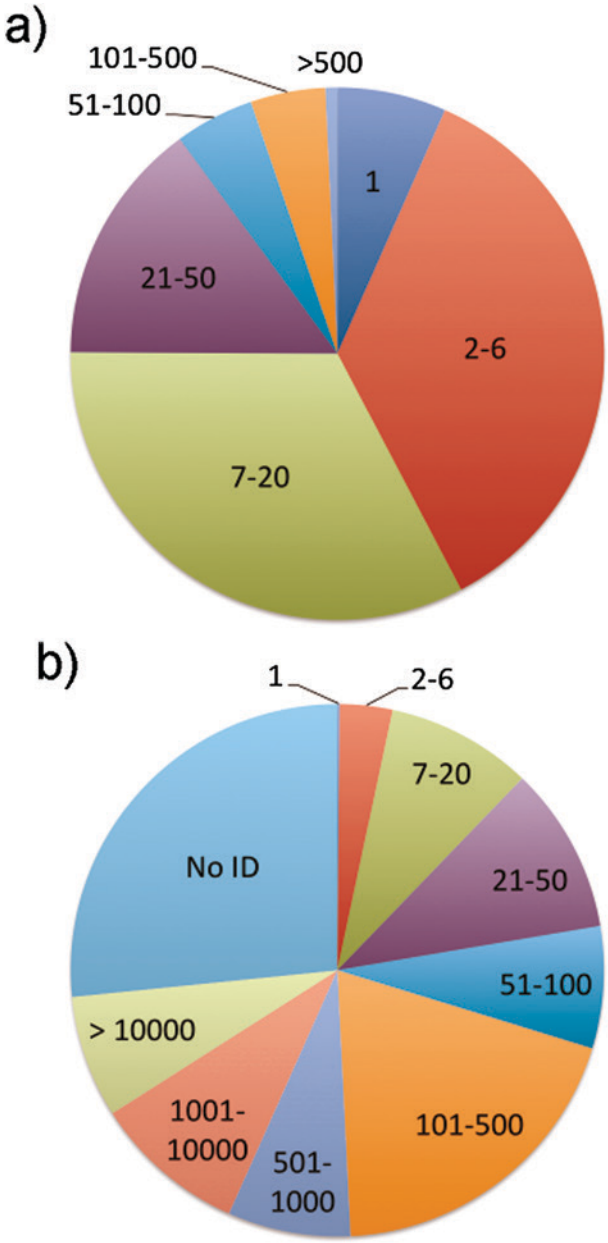


Fig. 1. (a) Breakdown of number of classifications per user and (b) percentage of total contributions from each tier of activity.

DATA AND DESIGN. Cyclone Center was designed to follow the principles of the enhanced infrared (EIR) version of the Dvorak technique (Dvorak 1984; Velden et al. 2006) in a way that citizen scientists could provide the required information for the intensity estimate. In general, the technique involves the interpretation of an IR image of a TC and the application of a number of constraints and set procedures. Dvorak has been used consistently at all global TC forecast agencies for at least the last 20 years, and validation studies (e.g., Gaby et al. 1980; Knaff et al. 2010) have shown that average differences between Dvorak intensity estimates and aircraft reconnaissance-based best-track data are quite low, ranging from 1.5 to 9 knots (kt; 1 kt = 0.51 m s^{-1}). In the absence of reconnaissance and/or other irregularly available data (e.g., scatterometers, microwave), the Dvorak technique still unquestionably provides the best estimate of TC intensity.

The IR satellite dataset used for all classifications is the Hurricane Satellite (HURSAT)-B1 archive (Knapp and Kossin 2007). HURSAT provides nearly 300,000 TC-centric images of 3,321 global tropical cyclones that formed from 1978 to 2009. The data were created by merging global geostationary satellite data and calibrating for homogenization; this allows us to classify global TC intensity across a 32-yr period without needing to account for differences in the observation platforms. However, using HURSAT for Cyclone Center does present a few challenges. The relatively coarse resolution (~8 km) can reduce the accuracy of the EIR technique.

Furthermore, classification skill would be optimized if other available HURSAT channels (e.g., visible, microwave) were used, but the visible channel is only available during daylight hours. In addition, regular microwave data are not available before 1997 and the relative scarcity of microwave passes over TCs creates substantial discontinuities in the time series [see Kossin et al. (2013) for more discussion]. To avoid the large logistical challenge in setting up the interface with discontinuous imagery, we use IR data exclusively, leaving the inclusion of other satellite data to future work.

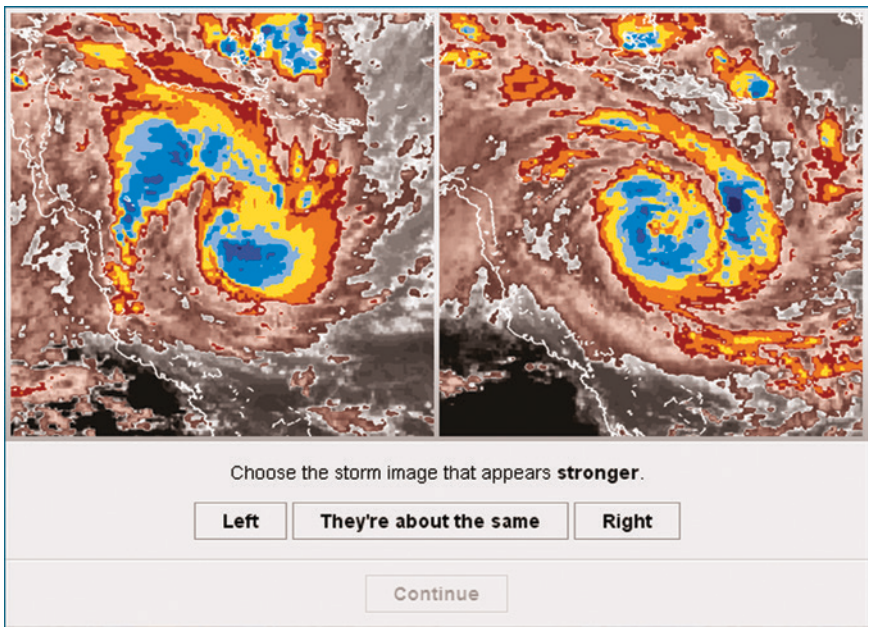


Fig. 2. Cyclone Center users are first asked to determine which image depicts a stronger storm. The image at right is the one which will be classified. The image at left is the storm 24 h earlier. Inline help (not shown) is provided to the users to guide them in their decision.

The Cyclone Center user interface was thus designed to gather data that are required to produce a classification using a modified EIR Dvorak method. However, we emphasize that the analysis presented here does not reflect the full EIR-like methodology, but rather the first steps, which involve determining the intensity trend and cloud pattern. Figure 2 is a screenshot that a user sees when they click on a storm to classify. Note the color enhancement used; the traditional Dvorak technique uses a repeating gray shade to distinguish between preestablished brightness temperature levels. To allow for easier user classifications, Cyclone Center uses a new, color-blind friendly enhancement scheme—see the sidebar on "Cyclone Center color enhancement" for more information. The first task asked of the user is to determine which image depicts the stronger storm. The responses are used to determine the intensity trend, which is needed to ultimately determine the current intensity.

The next step for the user is to choose a cloud pattern from one of the following: eye, embedded center (EMB), curved band (CBD), or sheared. A fifth pattern, called "other," is for storms that appear close to the satellite edge, extratropical, or do not exhibit any organized clouds at all (there is no allowance for subtropical or hybrid systems; classifiers proceed with the regular classification for these systems). As shown in Fig. 3, when a user selects a pattern that they perceive most closely matches the image, a subset of those patterns is shown and the user is instructed to choose the "closest match." These key images (Fig. 4) were selected from a pool of scene types with accompanying Dvorak evaluations, each done by a National Hurricane Center analyst during the 2004–06 Atlantic hurricane seasons. We selected images that we thought were the most representative of each respective cloud pattern and intensity; when the user selects an image, we have enough information to determine what is known as a "pattern T," or PT, number.

In the Dvorak technique, there are several kinds of "T numbers"; they are defined in Table 1. In the EIR Dvorak analysis, the PT number is an estimate of the TC's intensity based on the cloud pattern appearance and recent intensity trend (called the "MET"). Table 2 shows how EIR Dvorak T numbers are converted to the TC maximum sustained wind in the Atlantic and Pacific basins. In most operational cases, the PT number is not the value that is used as the final intensity estimate. It is preferred to determine the "data-T," or DT, number and then apply a number of rules and constraints, leading to the "final-T" (FT) number. Since we will ultimately calculate a DT-like value, there are questions presented to citizen scientists on the Cyclone Center page that provide the data we will need to do it. As mentioned previously, at this time we are focusing only on the questions that relate to the PT; the intent of this paper is to demonstrate that Cyclone Center is a viable way to achieve a skillful DT-like estimate of TC

intensity. We call our estimate of intensity the "pattern number" (PN), which diverges from PT as we consider the multiple classifications of an image.

For new or less experienced classifiers, there are resources available to help them through the process. As mentioned previously, an online tutorial introduces

CYCLONE CENTER COLOR ENHANCEMENT

The Dvorak method was originally developed using the "BD curve" enhancement for IR satellite imagery. This grayscale enhancement accommodated the technological limitations of the time. While experienced analysts can readily identify patterns in this enhancement, it can be confusing to the novice. Some shades are repeated, and it is not readily apparent which ones represent warmer or colder clouds. Since one of Cyclone Center's goals is to engage as many citizen scientists as possible, we developed a new enhancement scheme that more closely resembles traditional color scales with which the general public is familiar.

Our enhancement uses the same temperature thresholds as the BD curve, but the colors change from warm to cold with those temperatures. Both schemes use gray shading for values warmer than 9°C. The BD curve then uses a second series of grays, while we use a pink tint to add some differentiation. Both schemes use solid shades at varying intervals for temperatures colder than −30°C. Where the BD curve is forced to repeat two of its gray shades, our colorized scheme uses unique colors throughout.

Note in Fig. S1 that the BD curve uses black for temperatures from −63° to −69°C. This bold color marks a transition from moderate to tall clouds. This same transition is marked in our scheme by the change from reds and yellows to shades of blue. We also added an additional color (white) for temperatures colder than −85°C, which helps classifiers easily pick out the coldest cloud tops.

Since Cyclone Center seeks to maximize its pool of potential volunteers, we ensured that people with color vision deficiency could interpret our scheme. We were guided by the principles laid out by Light and Bartlein (2004). Specifically, we avoided schemes that included both red and green, and we sought a scheme that varied in both hue and intensity. Our ultimate selection was inspired by the "RdYlBu" scheme on http://colorbrewer2.org. The figure illustrates how people with the three most common color deficiencies would view our color enhancement.

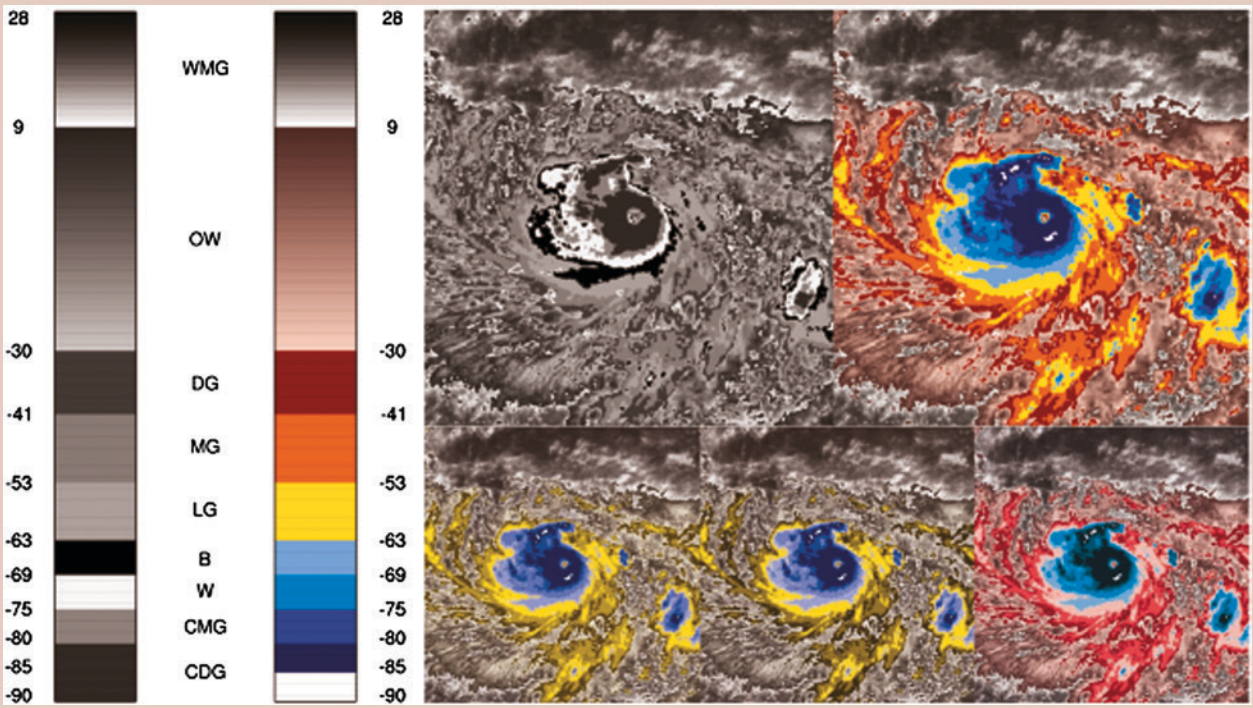


Fig. S1. Side-by-side comparison of Typhoon Gay (1992) using (clockwise from top left) original Dvorak BD enhancement, CC enhancement, and simulations of CC enhancement as seen with tritanopia, protanopia, and deuteranopia. Numbers are infrared brightness temperatures (°C) and Dvorak's named colors are identified in the middle: Warm Medium Gray (WMG), Off White (OW), Dark Gray (DG), Medium Gray (MG), Light Gray (LG), Black (B), White (W), Cold Medium Gray (CMG), and Cold Dark Gray (CDG). Simulations courtesy of http://vischeck.com.

classifiers to the interface and guides them through a complete classification. During "real" classifications, there is a dynamic help screen that is always shown on the right side of the browser, providing guidance in answering the current question. For example, classifiers are shown visual examples of what to look for (e.g., improved storm structure, colder cloud tops) to determine whether a storm had strengthened over the previous 24 h.

The Cyclone Center tool provides as output the selections of each individual, an individual's ID (if they log in), time classified, and image name. Each individual's identification is anonymous except for their personally provided ID. If an individual chooses not to log in, we still use their responses but they are treated a little differently in the weighting of the responses—this will be described in the following section.

To ensure that all storms in the database will be classified in a reasonable period of time, we retire an image when it acquires 10 unique classifications. This number gives us enough diversity to calculate a statistically reasonable consensus. A cost–benefit analysis (not shown) determined that the potential classification error was not sensitive to the number of classifications once 10 was reached (i.e., 30 classifications per image did not appreciably decrease the

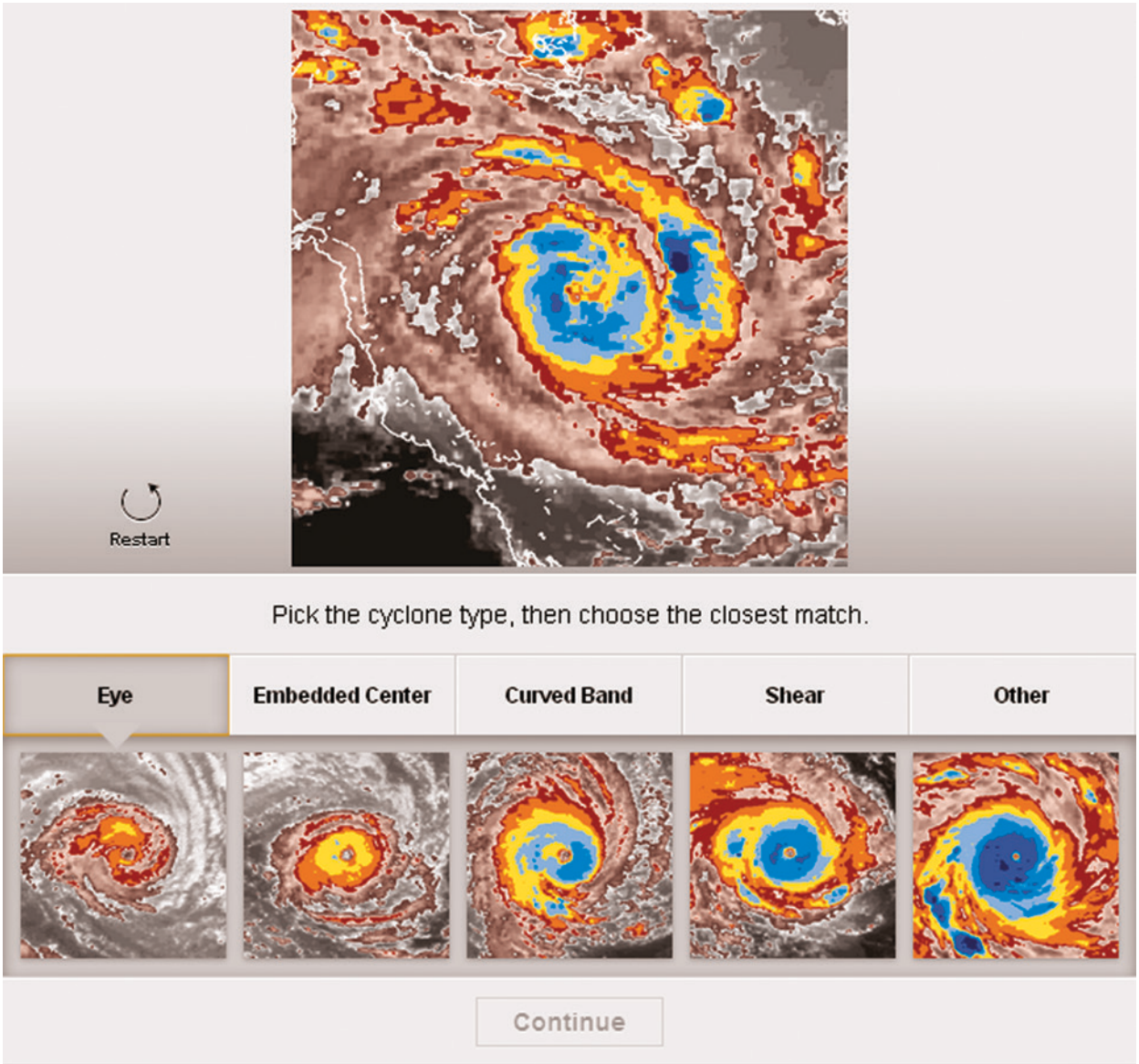


Fig. 3. After determining the intensity trend, users are asked to classify the cloud pattern and perceived organization of the TC. In this case, the user chose "eye" for the cloud pattern and is asked to choose the image that most closely matches the current one. Canonical images are oriented to match the storm's hemisphere.

Table 1. Numbers in the EIR Dvorak and Cyclone Center algorithms.		
Name	**Abbreviation**	**Description**
Model expected T	MET	TC intensity estimate based on comparison of TC structure and convective changes over the previous 24 h.
Pattern T	PT	TC intensity based on evaluation of the cloud pattern; cannot differ from MET significantly.
Data T	DT	Most objective methodology; uses a flow diagram for cloud pattern types and obtains TC intensity using brightness temperature criteria, distances, and other potentially subjective measures.
Final T	FT	Selected value of TC intensity as determined by the best choice of MET, PT, or DT; may be limited by established constraints.
Current intensity	CI	Final or official TC intensity of the Dvorak technique. Typically identical to FT, except during weakening phases of the TC when wind speed reductions typically lag the cloud patterns and trends.
Pattern number	PN	The PT-like value calculated from Cyclone Center data.

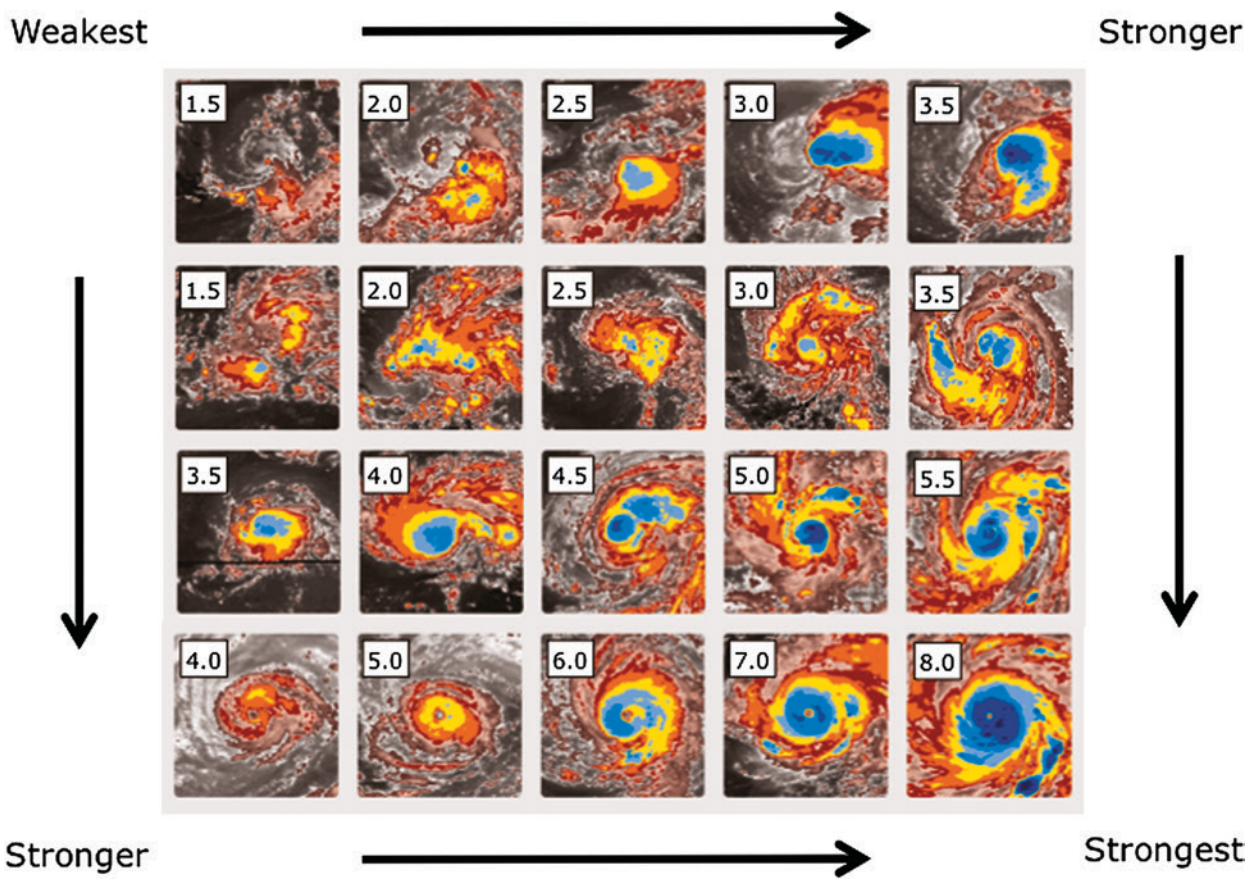


Fig. 4. Available cloud patterns from which classifiers can choose matches. The numbers represent the Dvorak data-T, or DT, number as determined by an analyst at the National Hurricane Center. Larger numbers correspond to stronger storms. Each row represents five intensity grades of a single cloud pattern. (top)–(bottom) The cloud patterns are shear, curved band, embedded center, and eye. Cloud patterns for the other options are not shown.

error). When all of the images for a particular storm are retired, the storm itself is taken out of circulation and replaced by a fresh one.

As of this writing, we have had over 8,000 unique users from around the world perform more than 400,000 classifications. All results presented here are from classifications recorded through 15 September 2013 and include approximately 5,000 unique users and 239,000 classifications.

DEVELOPING THE CONSENSUS. A significant challenge for this project is to estimate a storm's intensity at any time based on the numerous selections by citizen scientists. This can be complicated by the temporal dependence of the Dvorak technique—intensity change rates are limited based on storm type. Best estimates based on a crowd are not new (e.g., Brabham 2008); however, they may be new to the field of meteorology. Therefore, we employed two methods to initially look at the performance of the crowd toward estimating intensity from storm imagery. We demonstrate how a consensus approach allows the intercomparison of citizen scientists with subsequent spread being used to denote precision of an individual and estimate a bias-corrected intensity. We also use a Monte Carlo approach (section on "Consensus case studies") to randomly select individuals, which allows for the investigation of uncertainty in the resulting intensities.

In either case, there are two primary steps: 1) to estimate a given intensity based on a snapshot and 2) to apply temporal constraints to the instantaneous estimates. These initial calculations demonstrate a proof of concept rather than a complete analysis of a classification algorithm. To that end, we show the results of only a few storms and not a complete analysis of the technique.

The goal of the consensus approach is to combine each observation from all citizen scientists for a given snapshot into a PN that accounts for tendencies between each of them. The tendency of a given individual can be measured against observations from others when multiple citizen scientists view a common snapshot, as performed above. In this case, we are interested in consensus and will leave the calculations of best estimates to future work.

We currently focus on the pattern analysis from the citizen scientists. The input here is the image selected from the "Choose the closest match" question as shown in Fig. 3, chosen from all of the possible matches shown in Fig. 4. The calculation of a consensus PN (PN_i^c) for some snapshot in time i is calculated from all users k:

Table 2. EIR Dvorak T number/Cyclone Center PN conversion to maximum sustained wind. Note that maximum wind is assumed to be a 1-min average at a standard height of 10 m.

T number	Maximum wind (kt)
1.0	25
1.5	25
2.0	30
2.5	35
3.0	45
3.5	55
4.0	65
4.5	77
5.0	90
5.5	102
6.0	115
6.5	127
7.0	140
7.5	155
8.0	170

$$\mathrm{PN}_i^c = \frac{\sum_k w^k \left(\mathrm{PN}_i^k + B^k\right)}{\sum_k w^k}, \tag{1}$$

where PN_i^k is the classification from the kth classifier who is characterized by a weight w^k and a bias B^k. Hence, the consensus estimate is the weighted sum of the estimate from each available classification, where an individual's classification is corrected for their overall bias. Thus, it requires knowledge of a classifier's weight and bias, which are calculated as for user k by

$$w^k \propto \frac{\log N^k}{\sigma_k}, \quad B^k = \frac{\sum_i \left(\mathrm{PN}_i^k - \mathrm{PN}_i^c\right)}{N_m^k}, \quad \sigma_k^2 = \frac{\sum_i \left[\left(\mathrm{PN}_i^k - B^k\right) - \mathrm{PN}_i^c\right]^2}{N_m^k}, \tag{2}$$

where w^k is proportional to the number of classifications by the user N^k and inversely proportional to the mean deviation of the user from the consensus PN (σ_k). The bias and mean deviation are calculated

from a subset of N^k that have at least one classification from another individual N_m^k. The bias is the mean difference of a user's classifications from the consensus. Likewise, the mean deviation is the variance of a user's best estimate (corrected for bias) from the consensus. Future work will provide a more technical derivation of w^k and PN_i^c.

In this approach, the initial classifier characteristics are populated with random numbers. Then Eqs. (1) and (2) are iterated until the values of w^k and PN_i^c converge, usually after about four iterations. At this point, each image has a consensus intensity from available classifiers calculated from Eq. (1). Last, temporal constraints are applied to the individual values of PN_i^c following the advanced Dvorak technique [ADT; an automated, objective version of the Dvorak technique; see Olander and Velden (2007)], which provides an estimate of the final storm intensity. This value will be called the "CC consensus."

A limiting factor of this approach is the inability to identify characteristics of individuals who do not log in to the system. In this treatment, we apply a small weight and no bias to these individuals. The effect is that their classification is used when nobody else who is logged in has classified an image, whereas PN_i^c derives mostly from users that are logged in when possible, because of the small weight of those not logged in.

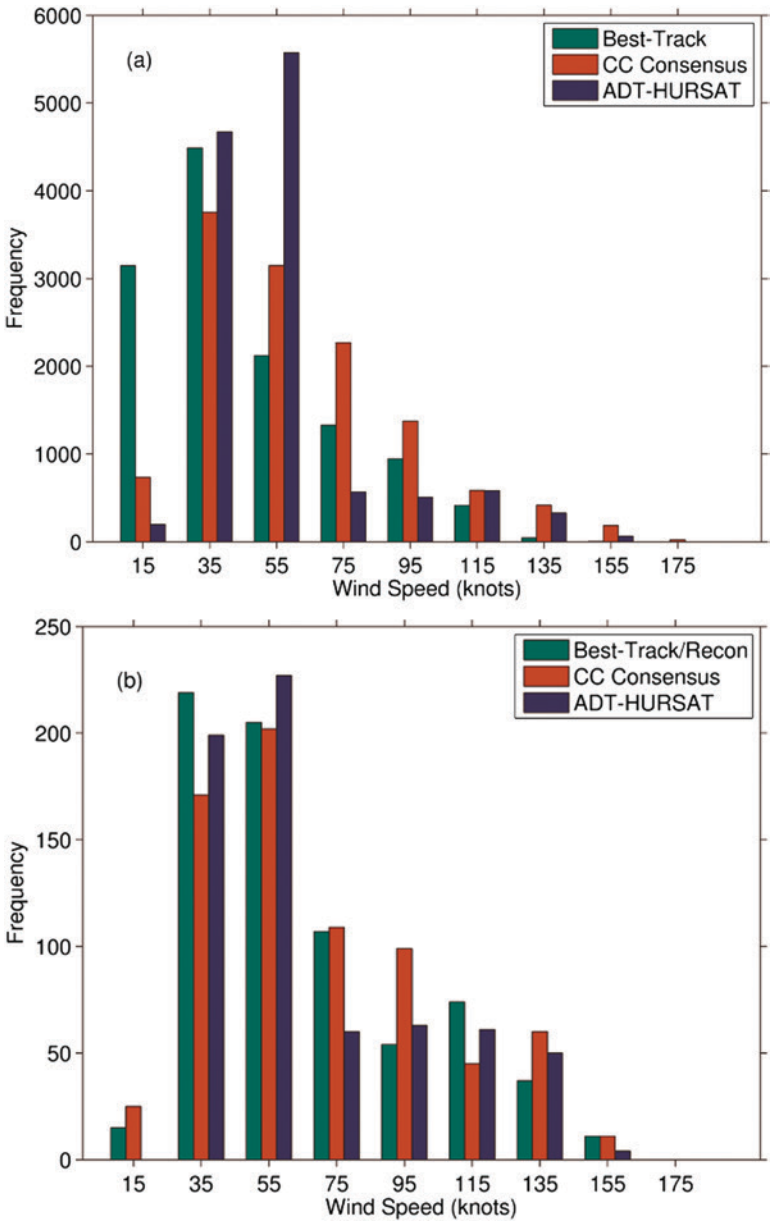


Fig. 5. Comparisons of intensity estimate distributions between Cyclone Center (CC consensus), ADT-HURSAT, and (a) global best track (IBTrACS), and (b) best track limited to times in the North Atlantic within 12 h of low-level aircraft reconnaissance (best-track/recon).

EARLY RESULTS. Proper evaluation of the performance of any algorithm that assesses TC intensity is difficult because the true intensity of a TC at any time is never exactly known. Even in cases where a storm is measured by multiple reconnaissance aircraft, the maximum surface wind is almost certainly missed because of the large ratio of storm size to observation area. For example, Uhlhorn and Nolan (2012) estimated that the maximum surface wind sampled by aircraft in a major (category 3 or higher) hurricane underestimates the true maximum wind by 7%–10%. We can however make a reasonable assessment of the CC performance through comparisons with existing TC intensity datasets (given all of their caveats). In this section we examine the value of the CC consensus with respect to best-track data (IBTrACS; Knapp et al. 2010) and an intensity dataset generated by the application of the ADT on HURSAT data (ADT-HURSAT; Kossin et al. 2013).

ADT-HURSAT comparisons. The ADT-HURSAT dataset was generated

from the same HURSAT imagery shown to citizen scientists on Cyclone Center, providing a convenient way to compare how humans classify images compared to a computer algorithm. ADT-HURSAT and CC consensus are compared to IBTrACS. Figure 5a shows the global wind speed distribution of IBTrACS ("best track"), CC consensus, and ADT-HURSAT for the 1978–2009 period. For the weakest best-track storms, both ADT-HURSAT and CC consensus tend to estimate higher intensities. The distribution of the CC consensus across all wind speeds appear to be more physically realistic than ADT-HURSAT. As discussed in Kossin et al. (2013), ADT-HURSAT experiences difficulty identifying changes in cloud pattern when a prehurricane intensity TC is transitioning from an embedded center to an eye. High cirrus clouds may linger over the developing eye and the ADT-HURSAT tends to hold onto the weaker pattern too long, resulting in an artificial frequency maximum centered at 55 kt [work is under way to improve the ADT in this regard—see Olander and Velden (2012)]. Our human classifiers appear to be better at identifying the pattern changes leading up to the emergence of an eye, as shown by the smoother transition to higher intensities in Fig. 5a; this will be better shown in a case study in the next section.

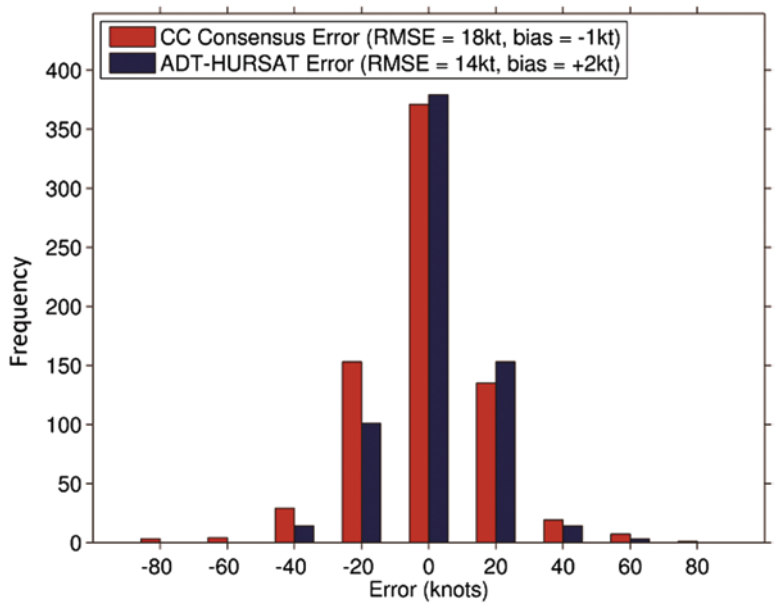


Fig. 6. Intensity estimate error distributions for CC consensus and ADT-HURSAT, as measured against best-track estimates within 12 h of low-level aircraft reconnaissance.

To focus on only the highest confidence "ground truth," further comparisons are made to a subset of North Atlantic TC points that are within 12 h of low-level aircraft reconnaissance. This subset, called "best-track/recon," contains 722 points that cover the 1995–2009 North Atlantic TC data record. Figure 5b shows the wind distributions for those cases. There is much more agreement between the three datasets for the weaker storms here, making it difficult to conclude whether the undercount from the weak storms highlighted in Fig. 5a is meaningful. Overall, Fig. 5b suggests both the CC consensus and ADT-HURSAT appear to do a good job at capturing the observed TC intensity distribution, though the ADT-HURSAT frequency minimum at the 75-kt bin remains.

Using the same best-track/recon validation set as ground truth, we calculated the root-mean-square error (RMSE) and bias for both ADT-HURSAT and the CC consensus (Fig. 6). Both datasets exhibit low bias and near-normal error distributions. The CC consensus RMSE is approximately 4 kt higher than ADT-HURSAT. The larger error is not surprising at this point. We expect that the CC errors will be reduced, perhaps substantially, when the images are subjected to a full EIR-like analysis. Also, there appear to be a number of egregious classifications in the CC dataset with errors exceeding 40 kt. A small number of these cases arise when classifiers incorrectly identify the TC of interest in a HURSAT image. Figure 7 shows one such case. Posttropical cyclone Nancy (0000 UTC 18 February 2005, near the image center) in the South Pacific is the intended classification image; here, Nancy's remnants appear to be sheared off to the southeast. Best-track data still issuing intensities on the storm list the maximum wind speeds at 30–35 kt. However, several CC classifiers incorrectly (but understandably) analyzed TC Olaf in the upper portion of the image. Olaf was clearly a mature TC at this point, with best-track winds of 100–120 kt (depending on agency). For this image, two classifiers identified Olaf as a PT = 7.0 eye pattern, one 5.5 embedded center, and one 2.0 shear case (ultimately thrown out by the consensus algorithm described in "Developing the consensus"). Three additional users analyzed a "no-storm" pattern for Nancy, producing a snapshot intensity (PN without temporal rules applied) of 6.4 (~125 kt). If best track is assumed ground truth, this produces an intensity error of approximately 90 kt. Although these cases are rare (~0.1% of all

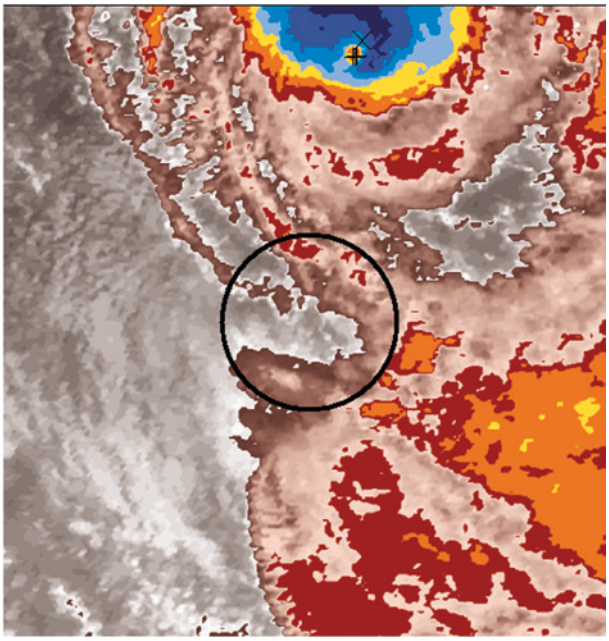


Fig. 7. An example of a misclassified storm. Nancy, whose center in IBTrACS is within the black circle, was the intended classification target. Four classifiers chose Typhoon Olaf at the top of the image to classify. Black crosses indicate user eye or embedded center fixes; the × shows a user center fix for a shear pattern.

images contain two storms, of which one is at least a hurricane and the other is a weak system), other cases similar to this one could explain some of the egregious errors seen in Fig. 7. Future work to identify user center fixes well off the image center should correct these cases.

Consensus case studies. Although descriptive statistics provide a good overview about the general performance of the CC consensus, we now present two case studies that provide more specific insight. Maximum wind speed time series of Typhoon Yvette (1992) and Typhoon Ivan (1997) are shown in Figs. 8a and 8c, respectively. In both of these storms, there is a large amount of disagreement between the best-track data of different forecast agencies (shown as gray lines). We recognize that a portion of the disagreements arise from several factors beyond Dvorak (pattern) interpretation, including different wind-averaging periods, inconsistent mapping of CI numbers to wind speeds, and other perceived regional bias adjustments. But recent work (Barcikowska et al. 2012; Nakazawa and Hoshino 2009) demonstrates that significant interagency intensity differences in operational Dvorak estimates drive disparate best-track data, even after accounting for these factors.

In Ivan (Fig. 8a), the CC consensus (green) closely follows the upper best-track data (JTWC) during the intensification and weakening phases but tends to agree more with other agencies in days 12–17. Yvette (Fig. 8c) displays more divergence between CC and the best-track data early on but closely follows the most intense best-track data (also JTWC) from day 9 onward. In both storms, the CC consensus nicely resolves the daily variance in the TC intensity and produces a maximum wind comparable to ADT-HURSAT.

As mentioned in the previous section, ADT-HURSAT has been shown to be sometimes late in identifying an eye pattern, resulting in an intensity plateau. This is clearly seen in Typhoon Ivan (Fig. 8a, magenta) on days 6–7, where ADT-HURSAT maintains an intensity of 60 kt while all other best-track and CC consensus shows an intensification trend [a similar plateau is seen in Yvette (Fig. 8c)]. A critical image in this scenario is shown in Fig. 9. At this point (1800 UTC 16 October 1997) a small, ragged eye is beginning to emerge from TC Ivan. ADT-HURSAT called this a central dense overcast (CDO, equivalent to the EIR embedded center) pattern and assigned a current intensity (CI) number of 3.9, just below typhoon strength (~63 kt). CC users were split, with 57% choosing embedded center and 43% eye pattern. The consensus PN is 5.7, analogous to a maximum wind speed of about 108 kt. Although the true intensity of Ivan at this point is arguable, our own manual EIR Dvorak analysis of this image assigns a DT of 6.0—equivalent to a 115-kt maximum sustained wind speed.

It is tempting to believe that these kinds of disagreements between ADT-HURSAT and the CC consensus permeate the dataset but this conclusion is not supported by the evidence. In fact, ADT-HURSAT and CC consensus (when unanimous) agree on certain cloud pattern types most of the time. When a TC is classified by at least five citizen scientists and they all agree on an eye cloud pattern, ADT-HURSAT concurs over 95% of the time. Similar results are seen with the embedded-center pattern. Other cloud patterns (e.g., shear and embedded center) show less agreement not only with ADT-HURSAT but also among CC classifiers themselves, suggesting they are less confident with those scene types. The sidebar presents an interesting analysis of CC classifier agreement on both cloud pattern type and overall intensity.

Monte Carlo approach. One limitation of the consensus approach is that it provides little information in the

way of how certain one can be about the intensity at any given time. While deviation between individuals can be calculated for each snapshot, it becomes convolved with the ADT temporal rules and information about uncertainty is lost.

A Monte Carlo approach can be used to estimate intensity and also address storm intensity uncertainty. We randomly select one classification from the available classifications for a snapshot. Performing this for each snapshot of the storm creates a simulated intensity analysis. Temporal rules (following ADT) are then applied to the random PN values, producing a final intensity time series of the system. However, there are numerous possible time series of PN based on differences between each citizen scientist. For instance, for a storm that lasts 7 days with 8 images per day and 10 citizen scientists per image, there are about ~10^{56} possible time series of PN (which is an upper limit given the likelihood that there would be some agreement in the classifications). In our analysis, we create 100 time series of PN through random selection of classifications at each time. This produces a distribution of intensities at each snapshot rather than one value. The variation of intensity at each snapshot provides an estimate of intensity and some measure of uncertainty.

The distribution of intensities using the Monte Carlo method is shown in Fig. 8b (Ivan) and Fig. 8d (Yvette). For both cases, the method shows that there is a large degree of agreement early on in the TC life cycles (days 1–5) and larger uncertainty during the mature stages of the TCs. We believe that the large uncertainty arises from the diversity of eye sizes, shapes, and eyewall cloud-top temperature patterns that may make it difficult to identify a close match on the eye pattern canonical images (row 4 in Fig. 4).

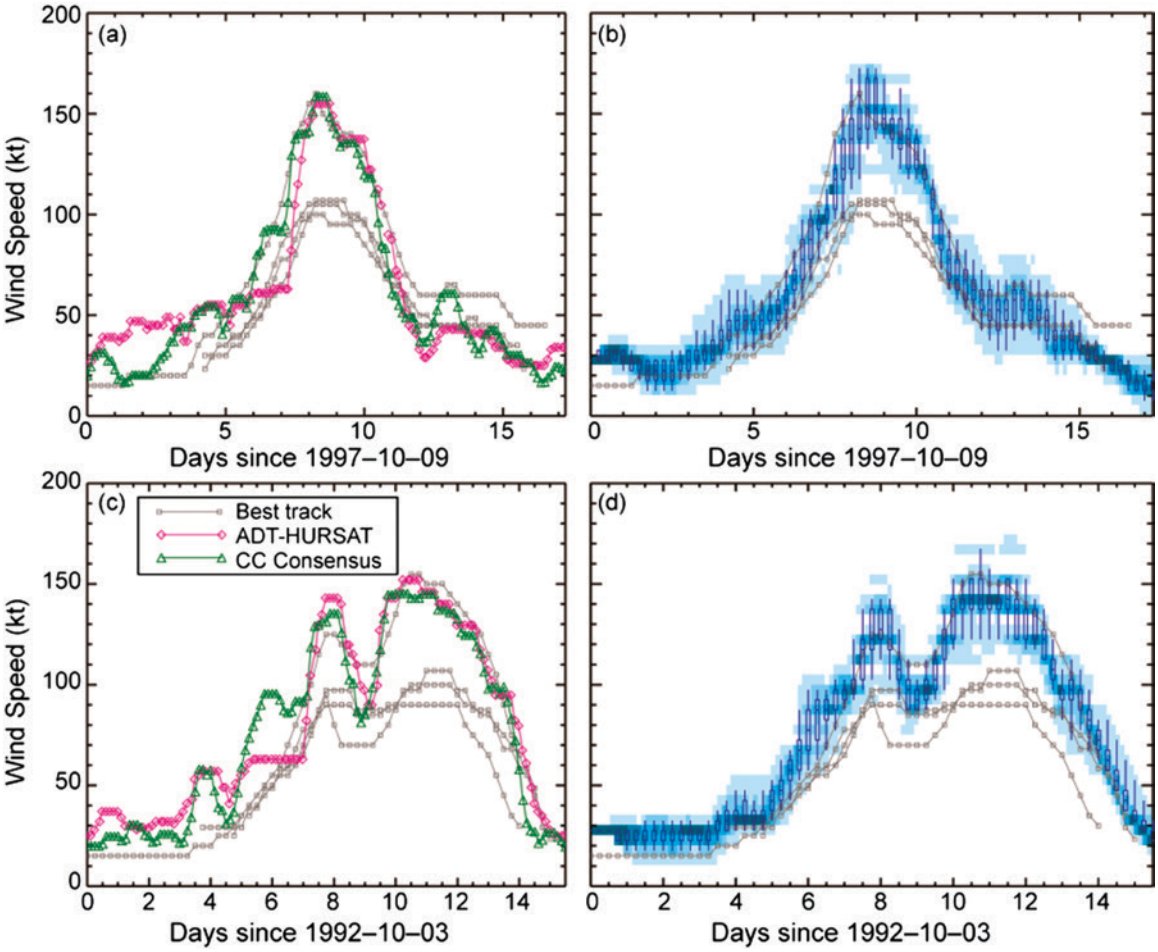


Fig. 8. (a) Time series of intensity of Typhoon Ivan (1997) with intensities from best-track data and estimates from ADT-HURSAT and Cyclone Center (CC consensus). (b) Time series of Ivan where shading is proportional to frequency of the simulated intensities (5-kt interval) and the box-and-whisker plots show the quartiles and the 5th and 95th percentiles (for clarity, box and whiskers are only shown with a 6-h interval). (c),(d) As in (a) and (b) (respectively), but for Typhoon Yvette (1992).

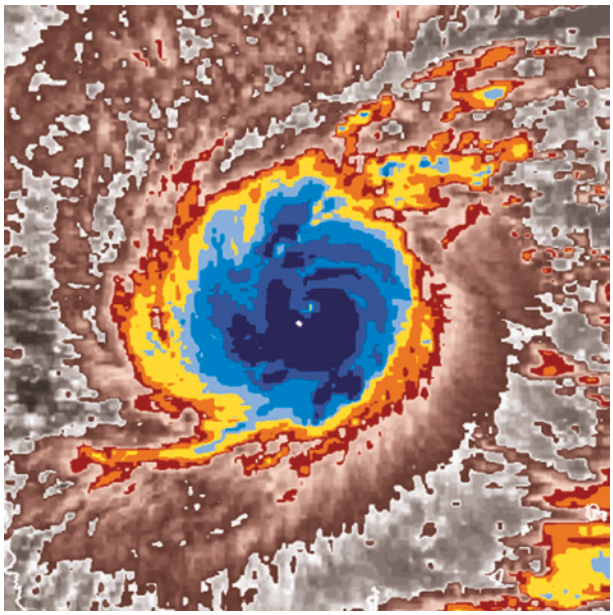


Fig. 9. Typhoon Ivan at 1800 UTC 16 Oct 1997. The yellow pixels surrounded by the dark blue near the center of the image represent the developing eye of the typhoon.

SUMMARY AND FUTURE WORK. Best-track data for TCs contain a high degree of uncertainty because TCs are rarely directly observed. Although the Dvorak technique is a ubiquitous and valuable tool for determining TC surface maximum wind speed, it is inherently subjective. Furthermore, global best-track data are also compromised by changes in technology such as improved satellite coverage and resolution, nonstandard changes to Dvorak rules and constraints at forecast agencies, and changes in the observation infrastructure. A global reanalysis of TCs is desirable and should be done. However, it is difficult to achieve without a large group of dedicated researchers and significant funding sources—especially if the participants have operational forecast and analysis commitments.

Cyclone Center employs a scheme that uses Dvorak-like pattern recognition on a 32-yr homogeneous satellite image dataset to provide consistent TC intensities. Cyclone Center is one of the first efforts to use crowd sourcing to analyze a large meteorological dataset. A website was developed to guide untrained users, called citizen scientists, to answer simple questions about TC cloud patterns and cloud temperatures. We have shown that

1) CC consensus intensity errors are comparable to ADT-HURSAT, even without the full EIR implementation;
2) CC classifications can be used to resolve gross discrepancies in best-track data; and
3) the crowdsourcing approach provides valuable information on uncertainty.

Our intention is not to modify or replace best-track data but rather to ultimately provide an objective assessment of modern TC intensity that may be used as a starting point for a future reanalysis project. Such a project could improve the CC consensus by including corrections for biases in TC intensity estimates that originate from the Dvorak technique, as demonstrated in Knaff et al. (2010).

The information presented here is just a hint of what is possible with the data that have been (and are still continuously being) collected. One valuable dataset that naturally falls out of a global TC survey is storm morphology information, such as storm size, eye size, eye temperature, eyewall temperature, and number of storms with strong banding features. These types of data can be easily extracted from the citizen scientist responses. The addition of the uncertainty information will provide an additional valuable piece of metadata that can aid analysts and researchers.

The implementation of the modified, complete EIR Dvorak procedure on the dataset is a high priority of the project going forward. As has been mentioned several times, the CC consensus results presented here are calculated from the "closest match" image selection combined with the TC intensity trend (e.g., the Dvorak PT number). This is a somewhat crude estimation of the TC intensity. We believe that the inclusion of additional pattern-specific information found in the EIR Dvorak technique will significantly improve our estimates of TC intensity. The inclusion of additional visible and microwave data is another step that we believe would make a significant improvement in the CC consensus. However, this would require a new development phase and launch of the website and is, therefore, reserved for another time.

Finally, we are working on a better way of creating the CC consensus. It is difficult to rate user skill level when there is no "gold standard" to measure against. Our current method weights users on the total number of classifications that they have done (more is assumed to be better) and their bias (how close they are to the consensus). Although this technique will minimize the effects of inexperienced and "crazy" classifiers, it does not fully take advantage of the highly skilled classifiers who may classify less but see the "right" pattern when others do not.

CLASSIFICATION CONSISTENCY: COMPARISONS TO A SUPERUSER

One method of evaluating Cyclone Center classifications is to compare an individual classification with other citizen scientists. In this case, we have selected to compare classifications with those from the most active Cyclone Center citizen scientist (and coauthor): the user "bretarn," who has 20,000+ classifications. So for a given snapshot image, how does bretarn classify snapshots compared to others? Figure S2 is a heat map representation showing the distribution of classifications from bretarn compared to all other individuals based on general storm pattern types: shear, eye, EMB, CBD, and other patterns (which includes posttropical, no storm, or on the satellite limb). The percentages represent the fraction of classifications from bretarn when another citizen scientist classifies the same image with a particular pattern type. For example, 53% of the time that any individual selected an eye storm, bretarn did, too. However, 35% of the time, an eye was confused with an embedded center. The largest agreement is for eyes and embedded centers. The shear case causes quite a bit of confusion, with bretarn agreeing only 15% of the time with other users. More often, when most other users select shear, bretarn selects embedded center (49%). Another category that has significant off-diagonal percentages is "other storms". However, this is expected since these storms are a catchall of categories. For example, a posttropical storm (other) can look like a shear pattern to the untrained eye. Similarly, one might classify a weak curved band as no storm (other) if they do not see a pattern in the cloud field. Yet, the impact of discrepancies in classifications is not as clear, since the intensities between the pattern types overlap. For instance, the intensity of a weak eye can be similar to a strong embedded center.

The impact of different pattern types is investigated quantitatively in Fig. S3. This heat map provides the same analysis as the first, except in terms of PN. The impact of selecting the wrong type is removed in this analysis and shows the agreement purely in terms of PN (where no-storm classifications were given a value of 0.1 and posttropical and limb storms were not included). The PNs show some agreement with some of the largest percentages lying on the diagonal; however, some patterns and outliers do occur. For example, bretarn classifies crowd-identified weak systems (PN = 0.1 or 1.5) as much stronger storms (PN = 3.5 or 4.0) 30% of the time. It cannot be determined at this point whether the systems are underrated by the crowd, overrated by bretarn, or a little of both; such an analysis is possible with the most active users. Tasks like these are the target of future work.

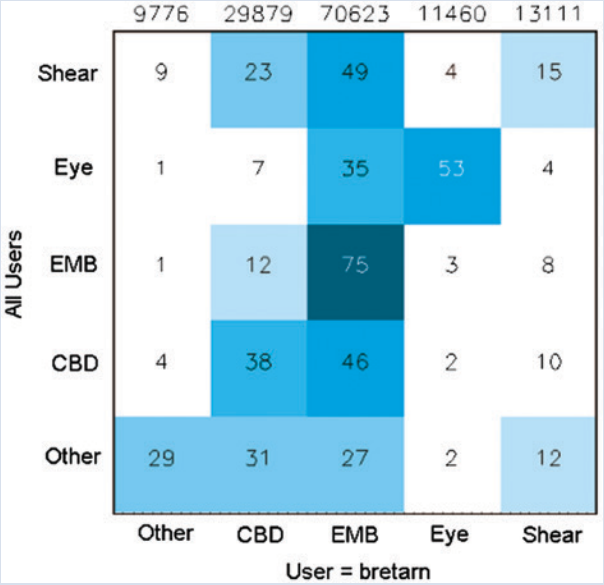


Fig. S2. Heat map of conditional frequency—the percentage of classifications from bretarn based on a given selection from all individuals classifying the same image as bretarn. Numbers along the top axis list the total number of snapshot matchups for each type from bretarn. Percentages are based on classifications from an individual, so rows sum to 100% (percentages rounded to integer).

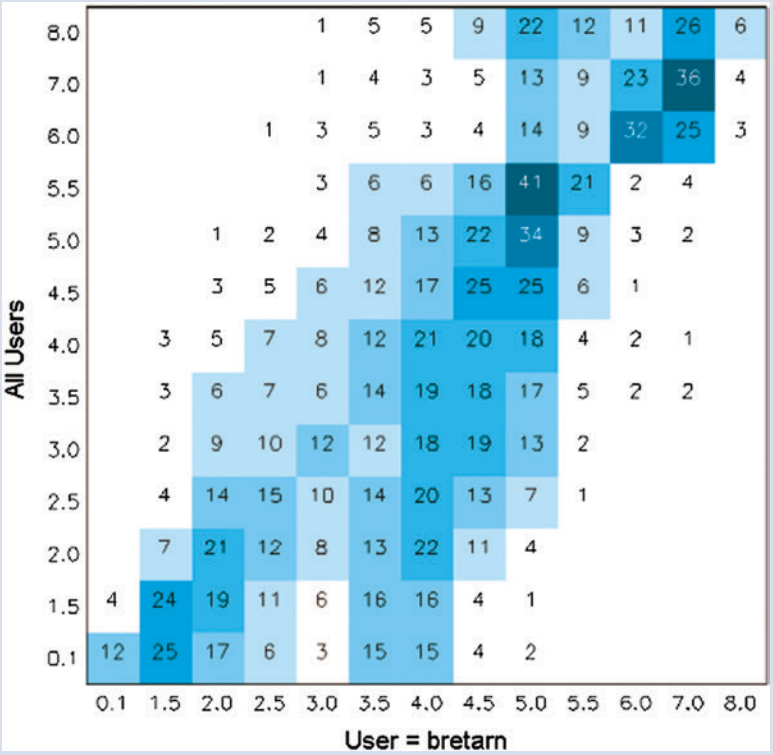


Fig. S3. Similar to Fig. S2, except in terms of PN. Each row is the distribution of bretarn selections based on a selection from citizen scientists.

Citizen scientists can help improve tropical cyclone records. Classifications on Cyclone Center (http://cyclonecenter.org) continue and the reader is encouraged to take part in this project.

ACKNOWLEDGMENTS. Cyclone Center was made possible through the support of a Citizen Science Alliance development grant from the Alfred P. Sloan Foundation and funding from the Risk Prediction Initiative (RPI). The team at Zooniverse, based at Adler Planetarium in Chicago, IL, worked closely with us on the design and development of the Cyclone Center website and continues to support the project to this day. In particular we acknowledge Brian Carstensen, Michael Parrish, Arfon Smith, Chris Snyder, David Weiner, Chris Lintott, David Miller, and Kelly Borden. We would like to also acknowledge the 8,000+ citizen scientists around the world who have made this research possible. Additional research support was provided by Brady Blackburn (Asheville High School/University of North Carolina at Chapel Hill) and Kyle Gayan (University of North Carolina at Asheville). Thorne was an employee of CICS-NC for the early portions of the project. Schreck and Stevens received support from the NOAA/Climate Data Record (CDR) Program through CICS-NC. Finally, we acknowledge the three thoughtful reviewers who provided valuable feedback and suggestions for future work.

REFERENCES

Barcikowska, M., F. Feser, and H. von Storch, 2012: Usability of best track data in climate statistics in the western North Pacific. *Mon. Wea. Rev.,* **140,** 2818–2830, doi:10.1175/MWR-D-11-00175.1.

Brabham, D., 2008: Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence,* **14,** 75–90, doi:10.1177/1354856507084420.

Cardamone, C., and Coauthors, 2009: Galaxy Zoo Green Peas: Discovery of a class of compact extremely star-forming galaxies. *Mon. Not. Roy. Astron. Soc.,* **399,** 1191–1205, doi:10.1111/j.1365-2966.2009.15383.x.

Diamond, H. J., A. M. Lorrey, K. R. Knapp, and D. H. Levinson, 2012: Development of an enhanced tropical cyclone tracks database for the southwest Pacific from 1840–2010. *Int. J. Climatol.,* **32,** 2240–2250.

Dvorak, V. F., 1973: A technique for the analysis and forecasting of tropical cyclone intensities from satellite pictures. NOAA Tech. Memo. NESS 45, 19 pp.

——, 1984: Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. 11, 45 pp.

Emanuel, K. A., 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature,* **436,** 686–688, doi:10.1038/nature03906.

Gaby, D. C., J. B. Lushine, B. M. Mayfield, S. C. Pearce, and F. E. Torres, 1980: Satellite classification of Atlantic tropical and subtropical cyclones: A review of eight years of classification at Miami. *Mon. Wea. Rev.,* **108,** 587–595, doi:10.1175/1520-0493(1980)1082.0.CO;2.

Gray, W. M., C. Neumann, and T. L. Tsui, 1991: Assessment of the role of aircraft reconnaissance on tropical cyclone analysis and forecasting. *Bull. Amer. Meteor. Soc.,* **72,** 1867–1883, doi:10.1175/1520-0477(1991)0722.0.CO;2.

Kantha, L., 2006: Time to replace the Saffir-Simpson hurricane scale? *Eos, Trans. Amer. Geophys. Union,* **87,** 3–6, doi:10.1029/2006EO010003.

Knaff, J. A., D. P. Brown, J. Courtney, G. M. Gallina, and J. L. Beven, 2010: An evaluation of Dvorak technique–based tropical cyclone intensity estimates. *Wea. Forecasting,* **25,** 1362–1379, doi:10.1175/2010WAF2222375.1.

Knapp, K. R., and J. P. Kossin, 2007: New global tropical cyclone data from ISCCP B1 geostationary satellite observations. *J. Appl. Remote Sens.,* **1,** 013505, doi:10.1117/1.2712816.

——, 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 best track data. *Bull. Amer. Meteor. Soc.,* **91,** 363–376, doi:10.1175/2009BAMS2755.1.

Knutson, T. R., and Coauthors, 2010: Tropical cylcones and climate change. *Nat. Geosci.,* **3,** 157–163, doi:10.1038/ngeo779.

Koba, H., T. Hagiwara, S. Osano, and S. Akashi, 1990: Relationship between the CI-number and central pressure and maximum wind speed in typhoons (in Japanese). *J. Meteor. Res.,* **42,** 59–67.

Kossin, J. P., K. R. Knapp, D. J. Vimont, R. J. Murname, and B. A. Harper, 2007: A globally consistent reanalysis of hurricane variability and trends. *Geophys. Res. Lett.,* **34,** L04815, doi:10.1029/2006GL028836.

——, 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.

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

——, and Coauthors, 2004: A reanalysis of Hurricane Andrew's intensity. *Bull. Amer. Meteor. Soc.,* **85,** 1699–1712, doi:10.1175/BAMS-85-11-1699.

——, and Coauthors, 2008: A reanalysis of the 1911–20 Atlantic hurricane database. *J. Climate,* **21,** 2138–2166, doi:10.1175/2007JCLI1119.1.

Light, A., and P. J. Bartlein, 2004: The end of the rainbow? Color schemes for improved data graphics. *Eos, Trans. Amer. Geophys. Union,* **85,** 385–391, doi:10.1029/2004EO400002.

Lintott, C. J., and Coauthors, 2008: Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Mon. Not. Roy. Astron. Soc.,* **389,** 1179–1189, doi:10.1111/j.1365-2966.2008.13689.x.

Martin, J. D., and W. M. Gray, 1993: Tropical cyclone observation and forecasting with and without aircraft reconnaissance. *Wea. Forecasting,* **8,** 519–532, doi:10.1175/1520-0434(1993)0082.0.CO;2.

Nakazawa, T., and S. Hoshino, 2009: Intercomparison of Dvorak parameters in the tropical cyclone datasets over the western North Pacific. *Sola,* **5,** 33–36, doi:10.2151/sola.2009-009.

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.

——, and ——, 2012: The current status of the UW-CIMMS Advanced Dvorak Technique. Preprints, *30th Conf. on Hurricanes and Tropical Meteorology,* Ponte Verde Beach, FL, Amer. Meteor. Soc., 7C.1. [Available online at https://ams.confex.com/ams/30Hurricane/webprogram/Paper204529.html.]

Poroseva, S. V., N. Lay, and M. Y. Hussaini, 2010: Multimodel approach based on evidence theory for forecasting tropical cyclone tracks. *Mon. Wea. Rev.,* **138,** 405–420, doi:10.1175/2009MWR2733.1.

Rappaport, E. N., and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting,* **24,** 395–419, doi:10.1175/2008WAF2222128.1.

Schenkel, B. A., and R. E. Hart, 2012: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *J. Climate,* **25,** 3453–3475, doi:10.1175/2011JCLI4208.1.

Schreck, C. J., K. R. Knapp, and J. P. Kossin, 2014: The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS. *Mon. Wea. Rev.,* **142,** 3881–3899, doi:10.1175/MWR-D-14-00021.1.

Seneviratne, S. I., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation,* C. B. Field et al., Eds., Cambridge University Press, 109–230.

Shoemaker, D. N., W. M. Gray, and J. D. Sheaffer, 1990: Influence of synoptic track aircraft reconnaissance on JTWC tropical cyclone track forecast errors. *Wea. Forecasting,* **5,** 503–507, doi:10.1175/1520-0434(1990)0052.0.CO;2.

Sievers, O., K. Fraedrich, and C. C. Raible, 2000: Self-adapting ensemble predictions of tropical cyclone tracks. *Wea. Forecasting,* **15,** 623–629, doi:10.1175/1520-0434(2000)0152.0.CO;2.

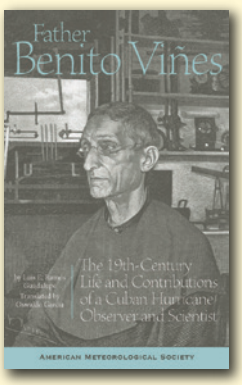
Uhlhorn, E. W., and D. S. Nolan, 2012: Observational undersampling in tropical cyclones and implications for estimated intensity. Mon. Wea. Rev., 140, 825-840, doi:10.1175/MWR-D-11-00073.1.

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.

Webster, P. J., G. J. Holland, J. A. Curry, and H.-R. Chang, 2005: Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science,* **309,** 1844–1846, doi:10.1126/science.1116448.

Wu, M.-C., K.-H. Yeung, and W.-L. Chang, 2006: Trends in western North Pacific tropical cyclone intensity. *Eos, Trans. Amer. Geophys. Union,* **87,** 537–538, doi:10.1029/2006EO480001.

NEW FROM AMS BOOKS!



FATHER BENITO VIÑES
The 19th-Century Life and Contributions of a Cuban Hurricane Observer and Scientist

BY LUIS E. RAMOS,
TRANSLATED BY OSWALDO GARCIA

Before Doppler radar, storm trackers, and emergency alerts, Father Benito Viñes (the "Hurricane Priest") developed the first network of weather observation stations in the Caribbean. His research at Belen Observatory in colonial Cuba laid the groundwork for present-day hurricane warning systems and kept people safer.

This biography portrays a pioneering citizen scientist who remained devoted to his religious life and includes notes from the translator that put his life into modern context.

© 2014, PAPERBACK, 172 PAGES,
ISBN: 9781935704921
LIST $20 AMS MEMBER PRICE $16

AMS BOOKS
www.ametsoc.org/amsbookstore

AMS titles now available as eBooks at springer.com

AMS BOOKS

RESEARCH APPLICATIONS HISTORY

www.ametsoc.org/amsbookstore

Springer

Scan to see AMS eBook titles at springer.com

AMERICAN METEOROLOGICAL SOCIETY