An Objective Algorithm for Detecting and Tracking Tropical Cloud Clusters: Implications for Tropical Cyclogenesis Prediction

CHRISTOPHER C. HENNON
University of North Carolina at Asheville, Asheville, North Carolina

CHARLES N. HELMS
The Florida State University, Tallahassee, Florida

KENNETH R. KNAPP
NOAA/National Climatic Data Center, Asheville, North Carolina

AMANDA R. BOWEN
NOAA/National Weather Service, Melbourne, Florida

(Manuscript received 25 August 2010, in final form 22 December 2010)

ABSTRACT

An algorithm to detect and track global tropical cloud clusters (TCCs) is presented. TCCs are organized large areas of convection that form over warm tropical waters. TCCs are important because they are the “seedlings” that can evolve into tropical cyclones. A TCC satisfies the necessary condition of a “preexisting disturbance,” which provides the required latent heat release to drive the development of tropical cyclone circulations. The operational prediction of tropical cyclogenesis is poor because of weaknesses in the observational network and numerical models; thus, past studies have focused on identifying differences between “developing” (evolving into a tropical cyclone) and “nondeveloping” (failing to do so) TCCs in the global analysis fields to produce statistical forecasts of these events.

The algorithm presented here has been used to create a global dataset of all TCCs that formed from 1980 to 2008. Capitalizing on a global, Gridded Satellite (GridSat) infrared (IR) dataset, areas of persistent, intense convection are identified by analyzing characteristics of the IR brightness temperature ($T_b$) fields. Identified TCCs are tracked as they move around their ocean basin (or cross into others); variables such as TCC size, location, convective intensity, cloud-top height, development status (i.e., developing or nondeveloping), and a movement vector are recorded in Network Common Data Form (NetCDF). The algorithm can be adapted to near-real-time tracking of TCCs, which could be of great benefit to the tropical cyclone forecast community.

1. Introduction

For many reasons, accurate forecasts of tropical cyclone development (tropical cyclogenesis) remain elusive. Tropical cyclones form in areas where little (if any) in situ data are available; thus, numerical models are initialized almost exclusively with data retrieved from satellites. These data, with little exception, do not provide the necessary space and time resolution at the surface and other atmospheric levels to properly initialize the large areas of convection [tropical cloud clusters (TCCs)] that may develop into a tropical cyclone.

A further challenge includes the physical modeling of important tropical cyclogenesis processes that are thought to operate on multiple scales. While several critical field experiments, like the Tropical Cloud Systems and Processes Experiment (TCSP; Halverson et al. 2007) and the National Aeronautics and Space Administration (NASA) African Monsoon Multidisciplinary Analysis (AMMA; Zipser et al. 2009), have provided important observational datasets that are being used to advance tropical
cyclogenesis theory, the numerical models have been slow to realize any significant increases in forecast skill. For example, Pratt and Evans (2009) examined the ability of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model to predict tropical cyclogenesis during the 2002–03 Atlantic hurricane seasons. They determined a forecast a “success” if the GFS predicted a closed isobar system within ±24 h in time and a 690-km radius in space of an actual tropical cyclogenesis event, a somewhat generous allowance. Their results showed a tendency for the GFS to overforecast genesis events. False alarm rates were ~33% for each season, while the Heidke skill scores [which range between 0 and 1, with higher values indicating more skillful forecasts; see Wilks (1995)] were 0.34 (in 2002) and 0.36 (in 2003). For comparison, purely statistical tropical cyclogenesis forecasts using large-scale data by Hennon et al. (2005) and Kerns and Zipser (2009) produced short-term (24 h) Heidke scores of ~0.40, indicating that recent statistical techniques have shown more skill at tropical cyclogenesis prediction than a recent version of the GFS.

However, developing and applying statistical tropical cyclogenesis models presents some fundamental challenges. First, statistical forecasts need to have a large number of past cases to be successful, which requires a large amount of labor to create. For example, Hennon et al. (2005) analyzed 4 yr (1998–2001) of Atlantic TCC activity by tracking TCCs manually in Geostationary Operational Environmental Satellite (GOES) infrared (IR) imagery, a process that took over 1 yr. More recently, Gierach et al. (2007) tracked TCCs using GOES imagery and noted that the process was time consuming, such that the number of analyzed tracks was limited. Second, the statistical models, especially those developed for operational forecasting, are usually reliant on the global model analysis fields for their predictor data. Thus, in a way, they are at the mercy of the model’s data and assimilation scheme.

The algorithm described here was developed to produce a high-quality TCC dataset that can, among other things, be used to improve tropical cyclogenesis forecasts. The algorithm uses a global Gridded Satellite (GridSat) dataset (see section 2a) that has been calibrated to produce consistent global IR brightness temperatures ($T_b$) over a 30-yr period (1980–2009); the results described here span 29 yr (1980–2008).

### a. Defining a tropical cloud cluster

A TCC provides the “preexisting disturbance” that is required before tropical cyclogenesis can occur. It provides the latent heat that is necessary for driving the developing primary and secondary circulations.

Qualitatively, a TCC can be defined as a large, concentrated group of thunderstorms that either form or move over a tropical ocean basin. Leary and Houze (1979) viewed TCCs as a group of individual cumulonimbus towers connected by a common cirrus shield. Houze (1982) expanded on this definition by describing the following four stages of TCC development: early stage (isolated convective towers), mature stage (cirrus shield), weakening stage (stratiform precipitation dominates), and dissipating stage (precipitation ceases). Our algorithm focuses on the mature TCCs, because they are easier to track and they provide a necessary condition for tropical cyclogenesis. An example of a mature TCC is shown in Fig. 1.

The Global Atmospheric Research Project (GARP) Atlantic Tropical Experiment (GATE) produced in situ observations of several hundred TCCs. Martin and Schreiner (1981) examined 526 TCCs that moved through the GATE array and found that they had an average lifetime of 28 h and an average size of $2 \times 10^5$ km$^2$, corresponding to a radius of ~250 km. A more recent study (Machado and Rossow 1993) found that about 20% (80%) of the total area of a TCC is convective (stratiform) in nature. Other mean properties of cloud clusters have been described by Ruprecht and Gray (1976a,b). For the development of the algorithm, it was necessary to further quantify these guidelines. For example, what should be used as a $T_b$ threshold for identifying sufficiently high cloud tops? How large does the TCC have to be? How long does it have to persist? The answers to these questions have varied considerably in other studies, and will be discussed in section 3a.
b. Previous TCC detection methods

Other automated algorithms for tracking convection have been developed and applied. Our relatively simple technique for tracking TCCs has a similar framework to an algorithm presented by Morel and Senesi (2002) and applied by Vila et al. (2008) to track (and ultimately forecast) mesoscale convection. This “area overlap method” initially assumes that a convective area [in this case a mesoscale convective system (MCS)] in a later scene is the same convection at a previous time, as long as it is within a previously defined distance.

More sophisticated automated techniques have also been applied to tracking meteorological phenomena. Lakshmanan et al. (2009) mention that the choice of a single threshold value for identification is a problem because it is impossible to distinguish between robust persistent convection and noise. Therefore, they apply an adapted watershed transform, an efficient technique for testing all possible thresholds in an image. Although we do not apply the watershed transform here, we do allow the threshold temperature to vary by ocean basin and limit noise by requiring strong temporal persistence in each TCC.

With regards to TCCs specifically, most tracking methods have been done manually. McBride (1981) summarized a large dataset of TCCs that were compiled at Colorado State University. The tracking methods varied by basin, depending on the type and amount of data available. In the Pacific basin, IR satellite data were used to find the center of the cirrus shield. In the Atlantic, the visible channel was included along with IR to find the center of cloud area mass. When applicable, McBride identified easterly waves to track “wave trough” clusters.

Perrone and Lowe (1986) used a similar manual technique to create a dataset of TCCs that formed from 1974 to 1977. They used a combination of National Oceanic and Atmospheric Administration (NOAA) tropical mosaic visible satellite images and postseason Joint Typhoon Warning Center (JTWC) best-track analyses to identify 130 developing (formed a tropical storm) and 460 non-developing (did not form a tropical storm) TCCs in the western North Pacific.

A similar “developing versus nondeveloping” TCC dataset was created by Hennon and Hobgood (2003) for the Atlantic basin. TCCs were subjectively identified over a 3-yr period (1998–2000) using a mosaic of 6-hourly GOES and Meteorological Satellite (Meteosat) IR imagery. Nearly 2200 scenes were available for TCC tracking during the Atlantic hurricane seasons; the 2-yr tracking process resulted in the identification of 291 TCCs; an additional year of TCCs was subsequently added to the dataset for a follow-up study (Hennon et al. 2005).

More recently, Kerns et al. (2008) created a different type of TCC dataset (1998–2001) for their tropical cyclogenesis study. They preface the description of their method by noting that TCCs and tropical waves are not always spatially collocated; this is an important point, because tropical waves provide favorable background vorticity that works to concentrate the latent heat produced by TCCs. Thus, it is desirable to identify TCCs that are associated with the larger background relative vorticity that tropical waves provide. Kerns et al. use 6-hourly 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40; Uppala et al. 2005) data for the Atlantic and eastern Pacific basins to identify maxima in the low-level relative vorticity field. Realizing that many of the identified tropical waves did not contain significant convection, Kerns and Zipser (2009) removed TCC candidates that did not have at least 10% coverage of cloud tops below 270 K within a 7° radius of the vorticity center. They argued that this allowed for realistic tropical cyclogenesis studies, because at least minimal convection was present in all of the final candidates. However, the 270-K threshold is much warmer than other thresholds for deep tropical convection; this will be discussed further in section 3a.

To our knowledge, the datasets described above have not been updated since their publication. Thus, there is no routinely updated dataset to support tropical cyclogenesis studies. The objective and automated nature of this algorithm will produce such a dataset. The algorithm’s use of merged and calibrated GridSat data allows for TCC tracking on a global scale back to 1980, producing the largest TCC dataset ever created; the GridSat data are described below in section 2. Section 3 describes the methodology of the algorithm. This is followed in section 4 by a preliminary look at the TCC dataset that was produced. Finally, conclusions and opportunities for future work are presented in section 5.

2. Data

The algorithm uses two datasets to create the TCC tracks: GridSat (Knapp et al. 2011) and International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010). Each will be described in more detail below.

a. GridSat

GridSat is a near-global (70°N–70°S) satellite dataset derived from the International Satellite Cloud Climatology Project (ISCCP) B1 data (Knapp 2008a). Because it
is global and spans many years, GridSat is composed of several different satellite platforms, including GOES (1–13), Meteosat (2–9), Geostationary Meteorological Satellite (GMS; 1–5), and several others. In total, 32 satellite platforms contribute to the dataset. Despite the large number of data sources, GridSat provides a homogenous set of IR observations that can be used for this algorithm and many other applications. This was accomplished by numerous calibration efforts, including those from ISCCP (Desormeaux et al. 1993) and Knapp (2008b). Essentially, GridSat can be thought of as a satellite reanalysis. Data are provided in 3-hourly intervals, which is enough temporal resolution to accurately track TCCs. See Knapp et al. (2011) for a complete description of the GridSat data.

b. IBTrACS

IBTrACS is a global best-track collection from the world’s tropical cyclone forecast centers. Best-track data usually result from a postseason analysis by the responsible forecast center, which uses any available data and information from both the operational and post-analysis phases. Generally, the result is a 6-hourly data file that contains the storm track, intensity, and often additional information (depending on the agency). IBTrACS contains best-track data from all World Meteorological Organization (WMO) Regional Specialized Meteorological Centers (RSMCs), as well as other national agencies (e.g., the Joint Typhoon Warning Center).

In some parts of the world, two or more agencies may produce a best track on the same storm. IBTrACS does not attempt to reconcile any differences between these tracks.

The IBTrACS data are used here to determine whether a TCC has developed into a tropical cyclone. The methodology describing how that is done is given in section 3f.

3. Methodology

This section describes the selection of the objective criteria for identifying a TCC and the process of applying the criteria to IR satellite scenes. As mentioned previously, there have been a wide range of criteria applied to identify TCCs. Perhaps the most useful guidance for objectively defining a TCC comes from Lee (1989). In his study of North Pacific basin TCCs, he adopted the following requirements for a TCC: it must be an independent entity, at least 2° radius in size, located no farther north than 40°, and persist for at least 24 h. These criteria have been adopted in a previous study (Hennon and Hobgood 2003) and provide a good starting point for this algorithm.

The algorithm was coded in the Interactive Data Language (IDL; see ITT 2010). IDL is data analysis and display software that is quite common in the sciences. Figure 2 shows a schematic diagram of the steps performed in the algorithm, each of which is discussed in more detail below.

**FIG. 2. Flowchart of the TCC algorithm.**

---

**Produce TCC Candidates**
- Get IR Satellite Data
- Apply Brightness Temperature Thresholds
- Remove areas of convection smaller than the minimum cluster size
- TCC candidates

**Process TCC Candidates**
- Remove TCC candidates located over land
- Check for independence
- Match TCC candidate to existing track or create a new track
- Check for persistence

**Finalize TCC Tracks**
- Interpolate missing points
- Get IBTrACS data and check for developing clusters
- Create final output
a. Size and intensity

The first step of the process is to examine a $T_b$ scene and determine which pixels meet the $T_b$ threshold for its ocean basin. The ocean basin boundaries, shown in Fig. 3, are based on the Hurricane Satellite (HURSAT) basin dataset (Knapp and Kossin 2007) but were modified to only include TCCs that formed between 30°N and 30°S. HURSAT basin is a precursor to the GridSat data.

Past experience has shown that tropical cyclones and other convection have diverse cloud-top temperatures in different basins resulting from differences in the atmosphere thermodynamic profiles and underlying ocean temperatures. Therefore, we allowed for the $T_b$ threshold to be a function of the ocean basin by sampling the $T_b$ data during each basin’s tropical cyclone “season” (the period during which most tropical cyclone activity occurs; see Table 1). If a basin had few tropical cyclones in its historical record (e.g., the South Atlantic), we sampled the local late summer–early autumn period. Two days were selected from each month, and two times (local day and night) were selected from each day. Distributions of North Atlantic and North Indian $T_b$ are shown in Fig. 4 as an example.

To choose the threshold value for each respective basin, we reviewed the literature for ranges of expected convective $T_b$ and then ran several sensitivity experiments with the algorithm. We compared the identified TCCs for each threshold experiment with the Hennon and Hobgood (2003) TCC dataset. For the North Atlantic basin, we found that a $T_b$ threshold of 224 K resulted in a dataset that most closely matched the output of that study (see section 4b for further comparisons to that study). That threshold is consistent with Machado and Rossow (1993), who found an average TCC $T_b$ of 221–229 K. The value of 224 K represents the top 2% coldest cloud pixels from the sample shown in Fig. 4. Therefore, we applied the 2% threshold for all other basins. The derived threshold temperatures are shown in Table 1.

All cloud pixels colder than the specified threshold are identified. Groups (adjacent pixels colder than the $T_b$ threshold) of acceptable pixels are then analyzed to determine a maximum radius from the geometric center of the group. This radius must be at least 1° [~111 km, which is identical to the criterion used by Perrone and Lowe (1986) in their study] in order for the group to be considered large enough to be a TCC. Furthermore, below-threshold cloud pixels must make up an area of at least 34 800 km² (~90% of the area of a 1° radius circle).

This filters out long, narrow cloud bands that may allow an irregularly shaped (noncircular) TCC to pass the size requirement. In summary, the algorithm only allows large, circular-like areas of continuous strong convection to be considered a potential TCC. These areas are passed on to the next step.

b. Location

A land mask is applied to eliminate TCCs over land. This is done to remain consistent with the motivation behind the creation of the algorithm, that is, to promote studies of tropical cyclogenesis, a process that cannot occur over land. Furthermore, initial experiments with the algorithm showed a number of identified TCCs that were not tropical at all. Several midlatitude systems, exhibiting baroclinic-like cloud features, were identified. Without a convenient method of using GridSat to distinguish between TCCs and baroclinic systems, we

<table>
<thead>
<tr>
<th>Basin</th>
<th>Threshold (K)</th>
<th>TC “season”</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Atlantic (NA)</td>
<td>224</td>
<td>1 June–30 November</td>
</tr>
<tr>
<td>South Atlantic (SA)</td>
<td>227</td>
<td>1 November–30 April</td>
</tr>
<tr>
<td>East Pacific (EP)</td>
<td>228</td>
<td>15 May–30 November</td>
</tr>
<tr>
<td>South Pacific (SP)</td>
<td>221</td>
<td>15 October–15 May</td>
</tr>
<tr>
<td>West Pacific (WP)</td>
<td>219</td>
<td>1 July–30 November</td>
</tr>
<tr>
<td>North Indian (NI)</td>
<td>218</td>
<td>1 April–15 December</td>
</tr>
<tr>
<td>South Indian (SI)</td>
<td>221</td>
<td>15 October–15 May</td>
</tr>
</tbody>
</table>
filtered out any system that formed poleward of 30° latitude. Although certainly not ideal, we believe that almost all tropical cyclogenesis events will still be captured in the data. However, it is still quite probable that the algorithm will miss a small number of TCCs that form poleward of 30°. It is hoped that a more intelligent way of distinguishing between cirrus and convective cloud tops can be installed in a future version. A promising approach would be to compare water vapor (WV) and IR brightness temperatures – but WV data are not available for the entire GridSat record.

c. Independence

TCCs can be complex systems with multiple and distinct areas of intense convection. It is common to find several TCC candidates within close proximity to one another. In these situations, the algorithm considers all of the convective maxima subsets of the largest (“parent”) cluster, as long as they are within 1200 km of the parent. If a candidate is more than 1200 km away, then it is considered a distinct TCC. The 1200-km threshold was found by running many experiments with the algorithm and comparing the output to the Hennon and Hobgood (2003) study.

d. Track

Once it is determined that a candidate passes all of the required criteria to be considered a TCC, the algorithm searches back in time to determine if the TCC existed previously. The geometric center of a TCC must be within a specified distance of its last recorded position in order to be identified as the same TCC. The distance depends on the number of hours since the last known fix and an assigned theoretical maximum TCC velocity. It is not required that TCCs pass all criteria described above for consecutive 3-h periods. In fact, TCCs can “disappear” for up to 12 h before regenerating somewhere within the specified search radius. This was established to allow for the natural (diurnal) pulses of strong convection that are common in TCCs; it is quite common for cloud tops over the tropical oceans to warm for a number of hours, especially during the daytime (e.g., Yang and Slingo 2001). The search radii for each time interval since the previous fix are given in Table 2. The radii were determined by first assuming an initial maximum TCC speed of 35 kt. After many experimental iterations, we then adjusted the search radii upward at earlier “search hours” (to account for large jumps in convective areas typical of TCCs) and downward at the 12-h search time. We found that the tuned search radii produced the most consistent tracks.

If the candidate TCC is determined to have existed at a previous time, the current location and all of the computed variables (see section 4a for details) are recorded in the data file. If the previous fix was more than 3 h from the current time, the location of the TCC at the missing times are linearly interpolated and all output variables are labeled as “missing.”

e. Persistence

The next step is to determine if the candidate TCCs persisted long enough to be considered valid TCCs. Persistent convection is required for tropical cyclogenesis; latent heat needs time to accumulate in the central area of
the disturbance. The algorithm checks the time duration of all of the identified TCC tracks and eliminates all that did not last for at least 24 h. The only exceptions are TCCs that formed into tropical cyclones. Given the difficulties in determining tropical cyclogenesis from satellite imagery, it is not uncommon for tropical cyclones to be declared after a relatively short, immature TCC phase.

f. Development status

The location of each TCC is compared in time and space to each tropical cyclone in the IBTrACS database. We apply a generous distance threshold (1000 km) to determine whether the TCC is near a tropical cyclone in IBTrACS; if it is, then the TCC is flagged as a “developing” cloud cluster and subsequent positions will not be recorded. Although the TCC time resolution is 3 h, the fixes are only checked with IBTrACS at the standard synoptic times (0000, 0600, 1200, and 1800 UTC). IBTrACS does not contain best-track information at other times; this is one of the reasons why the distance threshold is large.

It should be noted that there are several areas in the globe’s ocean basins where more than one forecast center issues track forecasts and best tracks. As mentioned previously, in these instances IBTrACS records each center’s points individually; there is no attempt to reconcile differences in either track or intensity. This algorithm makes no judgments regarding the most accurate track; if there is more than one fix in IBTrACS, then we average them together to yield one genesis point. Duplicate genesis points introduce additional uncertainty into the matching algorithm; this was another reason why a larger threshold value was chosen.

g. Track smoothing

It is quite common for convective centers of activity to “jump” from one section of the TCC to another during the TCC life cycle. The algorithm will detect these jumps, sometimes producing a quite chaotic track. Thus, the tracks that result from the algorithm do not represent the mean motion or translation speed, but rather a mean motion vector superimposed with anomalies caused by convective perturbations. Some users may be more interested in the mean; therefore, we apply a simple three-point running mean filter to the raw track to produce a smoothed one. The smoothed track is also available in the algorithm output.

4. Dataset characteristics

a. Calculated or recorded variables

When a TCC fix is made, the algorithm records numerous variables that are listed and briefly described in Table 3; most provide extended information on the size, shape, and convective intensity of the TCC. For instance, the algorithm calculates three different coordinates for the TCC center. The “geometric” center is the pixel closest to the center of all pixels that met the convective threshold. The “weighted” center is determined by spatially analyzing the $T_b$ field and nudging the center to be closer to the part of the TCC experiencing the most intense convection. Finally, the “convective” center identifies the center of the pixels with the coldest cloud tops. Figure 5 shows an example of the three different fixing techniques. Depending on the application of the data, one or more of these center locations may be useful.

As mentioned previously, when a TCC does not meet the criteria for an intermediate time (i.e., there was a wane in convection, or a “missed” fix), all of the coordinates, with the exception of the convective center, are estimated through a linear interpolation between the current and the last known fix. Also, all of the information about the convection characteristics of the “missed” TCCs is flagged as missing data in the output.

The cloud-top heights are determined by comparison of the pixel $T_b$ to an atmospheric sounding. The sounding for each TCC is derived from the McClatchey standard atmospheric profiles (McClatchey et al. 1972). The profiles are adjusted for latitude and time of year. The retrieved $T_b$ is matched to its profile and the cloud-top height is determined.

b. Preliminary results

1) COMPARISON TO A SUBJECTIVE ALGORITHM

A focused collocation study was done to provide useful information on the algorithm performance. TCC tracks were compared to tracks that were subjectively identified and described by Hennon and Hobgood (2003) for the 1999 North Atlantic season. One should not expect a high degree of agreement; although the identification criteria were similar, there were obvious biases in the subjective selection of TCCs. Nevertheless, useful information about the nature of the objective TCC tracks was discovered during the study.
Of the 16 TCCs that developed into at least tropical depressions, the objective algorithm matched 10 of those with the IBTrACS data (63%). This is similar to the number of matched developing TCCs globally (59%) over the 1982–2008 period (we excluded 1980–81 because of large amounts of missing data). There are a number of possible explanations for the “missing” developers, including missing satellite data (particularly in the Indian Ocean prior to 1998), missing higher-latitude TCCs, duplicate genesis fixes in IBTrACS (the algorithm uses the mean fix), storms that may have developed so quickly that they did not meet the algorithm requirements before being declared, and storms forming very close to land (may be masked out). Storms in the North Atlantic that are not matched by the algorithm in 1999 are Arelene, Tropical Depression (TD)-2, Bret, Emily, Jose, and Katrina.

The Hennon and Hobgood (2003) study identified 84 nondeveloping TCCs during the 1999 North Atlantic season. The objective algorithm tracked 157 nondeveloping TCCs during the same period (87% higher). To investigate this difference, the month of June was examined in more detail. Of the 19 nondeveloping TCCs identified subjectively, 9 had a direct match in the objective scheme. There were 26 TCCs in the algorithm that did not have a subjective match. The vast majority (21) of those cases was nearly stationary TCCs that were identified in the deep tropics within the intertropical convergence zone (ITCZ); these were cases that were not subjectively tracked because tropical cyclogenesis was so unlikely. Two TCCs were continuing tracks of subjective cases that were dropped, two were associated with a stalled frontal system off the east coast of the United States, and one was a central Caribbean TCC that appears promising but was not recorded in the subjective data.

This study shows that although the objective tracks produce a realistic and consistent TCC dataset, there are important concerns of which users must be aware. First, many developing cases will not be identified as such in the database for reasons mentioned above. Second, evidence suggests that the objective algorithm will identify a larger number of TCCs than subjective datasets. Third,
it appears that nontropical systems (e.g., stalled fronts) do appear in the data. Some quality control flags are included in the data and additional flags are planned to provide users with useful guidance on these topics and more. With these concerns in mind, we will now present some preliminary analyses of global TCC activity. We intend to publish a more detailed analysis at a later time.

2) GLOBAL TCC ACTIVITY

The tracking algorithm was run for the entire GridSat data period (1980–2009) with the exception of 2009 (2009 data will be made available). As mentioned previously, missing data during the 1980–81 seasons artificially lowered the number of tracked TCCs during that period. Furthermore, significant portions of the Indian Ocean were not observed until 1998, further lowering TCC numbers during the early portions of the data. A snapshot of the $T_b$ field for 1800 UTC 9 August 1999 with identified TCCs is shown in Fig. 6. Although an animation over several days is more effective at highlighting TCC patterns and movement, one can identify two TCCs in the North Indian Ocean, five in the west Pacific, one in the east Pacific, and one in the North Atlantic. Note the two very strong areas of convection in the east Pacific around 15°N; these are mature tropical cyclones [Hurricanes Dora (east) and Eugene (west)] that were filtered out by comparing their locations to the IBTrACS data. It is encouraging to note that three separate TCCs were identified in the broad area of convection in the far western Pacific; there was some concern that the algorithm would have difficulty in detecting cloud clusters in areas of very widespread convective activity.

Figure 7 is an illustration of the TCC density, or how many TCCs tracked within 55 km of a grid point (1998–2007) each year. In the North Atlantic, convection associated with easterly waves was regularly detected off the West African coast centered at 8°N, 20°W. Farther west, from the mid-Atlantic to eastern Caribbean, the concentration of TCCs is about half that in the eastern Atlantic. This result is consistent with past observations (Shieh and Colucci 2010), which have shown a climatological suppression of convection in this area. Note that there are few tracks in the South Atlantic because of the...
colder SSTs in the region. However, tropical cyclones have formed in that basin in the past, most notably Cyclone Catarina in March 2004.

The Pacific basin shows that many TCCs are associated with the west–east oriented ITCZ (≈8°N) and diagonally oriented South Pacific convergence zone (SPCZ; centered at ≈8°S, 160°E). Although most TCCs in these bands did not develop, tropical cyclogenesis is quite common, especially in the far eastern and western Pacific. There are also very dense TCC tracks in the Indian Ocean, which climatologically produces a number of tropical cyclones each year. Globally, Fig. 7 suggests that the algorithm is successful at capturing the expected spatial distribution of TCCs.

In total, there were over 45,000 TCCs identified from 1980 to 2008. Of these, almost 2400 eventually developed into tropical cyclones. These numbers dwarf the number of cases that have been used in other statistical studies of tropical cyclogenesis; a comparison of TCC datasets is shown in Table 4. On average nearly 1600 TCCs (≈83 of which developed into tropical depressions) were tracked each year. The numbers from the McBride (1981) dataset are low, probably because they had to rely solely on aircraft reconnaissance for several years (1961–65) of their data collection. McBride also restricted the time domain used in tracking. For example, TCCs in the northwest Pacific were tracked only during the months of June–September. Similarly, Kerns and Zipser (2009) restricted the domain of their study in both the eastern Pacific and North Atlantic. In any case, it is clear that the algorithm described in this paper provides by far the largest resource of TCC data.

c. Dataset format

To provide a large amount of adaptability to a diverse user community, the TCC dataset is now available in Network Common Data Form (NetCDF, Rew and Davis 1990) and comma-separated variable (CSV) format [available online at http://facstaff.unca.edu/chennon/research/tcc.shtml]. NetCDF was developed by Unidata, a division of the University Corporation for Atmospheric Research (UCAR). It is a data model for array-oriented scientific data access and is completely machine independent. NetCDF is popular in the atmospheric science community and is also widely used in other disciplines, such as oceanography and physics (Brown et al. 1993).

There are a number of advantages in using NetCDF beyond the machine-independent model. Data are easily shared among different applications, computing languages, and computer architectures through a common library. The inclusion of metadata is easily done and can vastly improve user interpretation and use of the data. Subsets of the dataset can be made efficiently. Furthermore, users do not have to change their own reading software if additional variables or metadata are added to the dataset at a future time.

5. Summary and conclusions

An algorithm is described that identifies and tracks tropical cloud clusters (TCCs) via global IR brightness temperature ($T_b$). Several criteria (described in section 3)

<table>
<thead>
<tr>
<th>Study</th>
<th>Domain</th>
<th>Years</th>
<th>Times (UTC)</th>
<th>TCCs per year/total TCCs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Northwest Atlantic</td>
<td>1967</td>
<td>0000 and 1200</td>
<td>20/157</td>
</tr>
<tr>
<td>Current study</td>
<td>Global</td>
<td>1980–2008</td>
<td>0000, 0300, 0600, 0900, 1200, 1500, 1800, 2100</td>
<td>1576/45 708</td>
</tr>
</tbody>
</table>
must be met in order for a TCC to be recorded as such. These criteria are size and intensity, independence, location, and persistence. Valid TCCs are recorded in a data file that will be available to users over the Internet.

a. Applications of TCC data

The development of this algorithm was motivated by two needs in the tropical cyclone community. First, prediction of tropical cyclone formation (tropical cyclogenesis) remains a difficult challenge primarily because of the paucity of observations over the global oceans and the absence of a development theory that can be realized in global numerical models. Current theories converge on the idea that there are critical processes or factors—whether they are vortex mergers (Simpson et al. 1997), critical latent–sensible heat fluxes (Emanuel et al. 1994), or inertial stiffening—that only become important if the synoptic-scale conditions are favorable for development. These factors operate (almost exclusively) at unobserved scales in the TCC. Thus, previous studies (e.g., Hennon and Hobgood 2003; Kerns and Zipser 2009) have focused on identifying large-scale differences between developing and nondeveloping cloud clusters. It takes a large amount of effort to compile case studies for these types of research; this algorithm was developed with this in mind. Now, thousands of cases will be immediately available for focused research on tropical cyclogenesis.

The second motivation was to create a dataset that can help to identify “missed” tropical cyclones in the global best-track data. It is accepted that there have been very few, if any, strong tropical cyclones that have escaped detection by forecast centers since the geostationary satellite era began in the 1960s. However, it is quite possible that there have been some weaker tropical cyclones (systems without a noticeable eye) that may not have been recorded. Adding these systems to global best-track files can mitigate some of the problems that others have found in using these data for climate change studies (Landsea 2007).

Given the long time series of TCC data, it will also be possible to analyze the TCC data for climatic trends. Although ideally one would like to have climate data span over several decades, the length of this dataset approaches the 30-yr climatological averaging period.

b. Future work

The nature of the algorithm lends itself to near-real-time processing of TCCs, assuming that IR \( T_b \) data can be obtained quickly. A simple web interface will be developed to allow access to data for TCCs of interest. At the end of the year, the operational data that were logged will be postprocessed (following the last two steps described in sections 3d–g) to augment the research dataset.

The algorithm will also be modified to ingest other TCC-centric atmospheric and oceanic data that may be of interest to the tropical cyclone community. These data will be extracted or computed from global reanalyses such as the NCEP–National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996) and will include variables such as vertical wind shear, integrated precipitable water, wind divergence, maximum potential intensity, and sea surface temperature.

Acknowledgments. This work was primarily supported by the NOAA Climate Change Data and Detection (CCDD) program Opportunity OAR-CPO-2009-2001430. Mr. Helms also received support from a UNC Asheville Undergraduate Research Program Advisory Council research grant. Discussions with Mark DeMaria inspired the development of this algorithm. We would also like to thank the IBTrACS team at the National Climatic Data Center for their helpful suggestions. Comments from three anonymous reviewers improved the manuscript.

REFERENCES


Knapp, K. R., 2008a: Scientific data stewardship of International
Satellite Cloud Climatology Project B1 global geostationary
——, 2008b: Calibration of long-term geostationary infrared
observations using HIRS. J. Atmos. Oceanic Technol., 25,
183–195.
——, and J. P. Kossin, 2007: New global tropical cyclone data from
ISCCP B1 geostationary satellite observations. J. Appl.
Remote Sens., 1, 013505.
——, M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann,
2010: The International Best Track Archive for climate stew-
ardship (IBTrACS). Bull. Amer. Meteor. Soc., 91,
363–376.
——, and Coauthors, 2011: Globally gridded satellite (GridSat)

Lakshmanan, V., K. Hondl, and R. Rabin, 2009: An efficient,
general-purpose technique for identifying storm cells in geo-
spatial images. J. Atmos. Oceanic Technol., 26,
523–537.

Landsea, C. W., 2007: Counting Atlantic tropical cyclones back to
1900. Eos, Trans. Amer. Geophys. Union, 88,

of convection in a tropical cloud cluster. J. Atmos. Sci., 36,
437–457.

Lee, C. S., 1989: Observational analysis of tropical cyclogenesis in
the western North Pacific. Part I: Structural evolution of cloud
clusters. J. Atmos. Sci., 46,
2580–2598.

Machado, L. A. T., and W. B. Rossow, 1993: Structural characteris-
tics and radiative properties of tropical cloud clusters.
Mon. Wea. Rev., 121,
3234–3260.

Martin, D. W., and A. J. Schreiner, 1981: Characteristics of West
African and East Atlantic cloud clusters: A survey from
GATE. Mon. Wea. Rev., 109,
1671–1688.

McBride, J., 1981: Observational analysis of tropical cyclone for-
mation: Part I: Basic description of data sets. J. Atmos. Sci., 38,
1117–1131.

McClatchey, R. A., R. W. Fenn, J. E. A. Selby, F. E. Volz, and J. S.
Garing, 1972: Optical properties of the atmosphere. Air Force

Morel, C., and S. Senesi, 2002: A climatology of mesoscale con-
vective systems over Europe using satellite infrared imagery: I:
Methodology. Quart. J. Roy. Meteor. Soc., 128,

Perrone, T. J., and P. R. Lowe, 1986: A statistically derived pre-
Rev., 114,
165–177.

Pratt, A. S., and J. L. Evans, 2009: Potential impacts of the Saharan
air layer on numerical model forecasts of North Atlantic
tropical cyclogenesis. Wea. Forecasting, 24,
420–435.

for scientific data access. Preprints, Sixth Int. Conf. on In-
teractive Information and Processing Systems for Meteorology,
Oceanography, and Hydrology, Anaheim, CA, Amer. Meteor.

Ruprecht, E., and W. M. Gray, 1976a: Analysis of satellite-observed
tropical cloud clusters. Part I: Wind and dynamic fields. Tellus,
28,
391–413.

——, and ———, 1976b: Analysis of satellite-observed tropical cloud
clusters. Part II: Thermal, moisture, and precipitation. Tellus,
28,
414–426.

Shieh, O. H., and S. J. Colucci, 2010: Local minimum of tropical
Soc., 91,
185–196.

Simpson, J. E., A. Ritchie, G. J. Holland, J. Halverson, and S. R.
Stewart, 1997: Mesoscale interactions in tropical cyclone
genesis. Mon. Wea. Rev., 125,
2643–2661.

Quart. J. Roy. Meteor. Soc., 131,
2961–3012.

Vila, D. A., L. Augusto, T. Machado, H. Laurent, and I. Velasco,
2008: Forecast and tracking the evolution of cloud clusters
(FoTrCC) using satellite infrared imagery: Methodology
and validation. Wea. Forecasting, 23,
233–245.


Yang, G.-Y., and J. Slingo, 2001: The diurnal cycle in the tropics.
Mon. Wea. Rev., 129,
784–801.

Zipser, E. J., and Coauthors, 2009: The Saharan air layer and the
fate of African easterly waves—NASA’s AMMA field study of
tropical cyclogenesis. Bull. Amer. Meteor. Soc., 90,
1137–1156.